Report No. NAWCADWAR-93031-60

The Adaptive Function allocation for Intelligent Cockpits (AFAIC) program:

Interim research and guidelines for the application of adaptive automation.

Jeffrey G. Morrison, Ph.D. (Ed.)
Human Factors Technology Development Branch (Code 6021)
Air Vehicle and Crew Systems Technology Department
Naval Air Warfare Center - Aircraft Division
P.O. Box 5152
Warminster, PA 18974-0591

23 May 1993

PROGRESS REPORT

Approved for Public Release; Distribution is Unlimited.

Prepared for:
Air Vehicle and Crew Systems Technology Department
Naval Air Warfare Center - Aircraft Division
P.O. Box 5152
Warminster, PA 18974-0591

9300130

NOTICES

REPORT NUMBERING SYSTEM — The numbering of technical project reports issued by the Naval Air Warfare Center, Aircraft Division, Warminster is arranged for specific identification purposes. Each number consists of the Center acronym, the calendar year in which the number was assigned, the sequence number of the report within the specific calendar year, and the official 2-digit correspondence code of the Functional Department responsible for the report. For example: Report No. NAWCADWAR-92001-60 indicates the first Center report for the year 1992 and prepared by the Air Vehicle and Crew Systems Technology Department. The numerical codes are as follows:

CODE	OFFICE OR DEPARTMENT
00	Commanding Officer, NAWCADWAR
01	Technical Director, NAWCADWAR
05	Computer Department
10	AntiSubmarine Warfare Systems Department
20	Tactical Air Systems Department
30	Warfare Systems Analysis Department
50	Mission Avionics Technology Department
60	Air Vehicle & Crew Systems Technology Department
70	Systems & Software Technology Department
80	Engineering Support Group
90	Test & Evaluation Group

PRODUCT ENDORSEMENT — The discussion or instructions concerning commercial products herein do not constitute an endorsement by the Government nor do they convey or imply the license or right to use such products.

Reviewed By: J. Schmidt LCDR MSC USN Date: 27SEP93
Branch Head

Reviewed By: S M Mitchell CDR, MSC, USN Date: 28SEP93
Division Head

Reviewed By: Donald P. McErlean Date: 28 Oct 93
Director/Deputy Director

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	23 May 93	Final Report

4. TITLE AND SUBTITLE
The Adaptive Function Allocation for Intelligent Cockpits (AFAIC) Program: Interim Research Guidelines for the Application of Adaptive Automation.

5. FUNDING NUMBERS
Program Element:
0602234N
Subproject:
RS34H20

6. ~~AUTHOR(S)~~
Jeffrey G. Morrison, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Human Factors Technology Branch (Code 6021)
Air Vehicle and Crew Systems Technology Department
Naval Air Warfare Center - Aircraft Division
P.O. Box 5152
Warminster, PA 18974-0591

8. PERFORMING ORGANIZATION REPORT NUMBER
NAWCADWAR-93031-60

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
See Block #7

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Report contains papers presented at the "Seventh International Symposium on Aviation Psychology" , held at Columbus, Ohio from 27-29 April 1993.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
This report encompasses seven papers presented at the Seventh International Symposium on Aviation Psychology on adaptive automation technology. Adaptive automation differ from conventional automation in that it is capable of invoking itself (i.e. turning itself on) with or without the explicit consent of a system operator and it may effec the nature of the task(s) performed by an operator in a system. Conceptually, the adaptive automation may affect the same task in different ways at different times, and need not necessarily assume full responsibility for a given task. Thus, adaptive automation is conceptually much more flexible in how it effects tasks and dynamic in the frequency with which it can change. Adaptive automation has the potential to solve many problems that are created and/or not addressed by conventional automation; however, it must be empirically studied and validated before this technology can be effectively applied to person-machine systems. The research described in this report presents contemporary research into a variety of adaptive automation issues: workload, situation awareness, cycles of automation, control and authority, and alternative interface concepts. In addition, a set of prospective, theoretically derived principles and guidelines are also presented.

14. SUBJECT TERMS
adaptive automation, workload, modeling, situation awareness, function allocation, human factors, display interfaces, person-machine interaction, strategies, guidelines, taxonomy.

15. NUMBER OF PAGES
48

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

FOREWORD

The papers incorporated in this report were written by researchers at the Catholic University of America, the University of Minnesota - Human Factors Research Laboratory, and the Naval Air Warfare Center - Aircraft Division - Warminster. The work was performed as part of the *Adaptive Function Allocation for Intelligent Cockpits* (AFAIC) program. This program is a 6.2 block-funded research program tasked with the development of human performance based principles and guidelines to help guide the application of Adaptive Automation technology to the tactical aircraft cockpit.

Adaptive Automation may be defined as: An approach to the automation in a person-machine system where the control of tasks may be by either the person or the machine and the machine has some degree of autonomy in changing the status or form the automation takes. In other words, the automation system may turn itself on or off, or change the nature of how the human operator is performing the task or tasks. By employing adaptive automation in a highly complex and demanding environment, such as that found in the tactical cockpit, it is expected that the pilot's workload can be kept at optimal levels, and the overall effectiveness of the pilot-vehicle system can be significantly improved.Major objectives of the AFAIC program include:

- Identify critical human performance issues in adaptive automation.
- Identify/develop methodologies and metrics appropriate to the study of human performance in adaptively automated systems.
- Perform research to explore the issues and validate human performance benefits of adaptive automation.
- Develop a set of prospective, theoretically derived principles and guidelines for application of adaptive automation technology to the tactical cockpit.
- Assess aircrew acceptance of adaptive automation in the crewstation.
- Disseminate data and lessons learned to research, engineering, and pilot-user communities.

The work described in this report describe some of the studies performed in order to meet these objectives during the third and fourth years of the program. As of the date this report is being written, the AFAIC program has been extended to continue into a fifth year. Additional technical reports detailing these and other studies are being prepared and/or planned, and a final summary report for the program will be summarize that work. The interested reader is encouraged to contact Jeffrey Morrison at NAWCAD-WAR for information regarding those reports.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank both the *Adaptive Function Allocation for Intelligent Cockpits* (AFAIC) program sponser, Mr. Jeff Grossman of the Naval Research and Development Center, and the Naval Air Warfare Center - Aircraft Division, Warminster, Code 60 management for their support of the AFAIC program. Their continued support of the program has contributed significantly to its success and is sincerely appreciated.

ABSTRACT

This report incorporates a series of seven papers presented at the *Seventh International Symposium on Aviation Psychology* as part of two sessions on adaptive automation technology. Adaptive automation differs from conventional (traditional or static) automation in two important aspects. First, is capable of invoking itself, i.e. turning itself on or off, with or without the explicit consent of the human in a person-machine system. Second, it may effect the nature of the task(s) performed by the human in a person-machine system in a number of ways. Conceptually, the automation may affect the same task in different ways at different times, and need not necessarily assume full responsibility for a task that it is effecting. Thus, adaptive automation is conceptually much more flexible in how it effects tasks, and dynamic in the frequency with which it can change status. These aspects of adaptive automation give it the potential to solve many problems that are created and/or not addressable with conventional automation, yet at the same time introduce a number of new human-machine interaction problems that must be empirically studied before this technology can be effectively applied to person-machine systems. The research described in the papers of this report describe contemporary research into a variety of issues in adaptive automation. These include: workload, situation awareness, cycles of automation, the issue of control and authority in adaptive automation, and alternative interface concepts for and adaptively automated crewstation. In addition, a set of prospective, theoretically derived principles and guidelines are presented.

TABLE OF CONTENTS

ABSTRACT

This report incorporates a series of seven papers presented at the *Seventh International Symposium on Aviation Psychology* as part of two sessions on adaptive automation technology. Adaptive automation differs from conventional (traditional or static) automation in two important aspects. First, is capable of invoking itself, i.e. turning itself on or off, with or without the explicit consent of the human in a person-machine system. Second, it may effect the nature of the task(s) performed by the human in a person-machine system in a number of ways. Conceptually, the automation may affect the same task in different ways at different times, and need not necessarily assume full responsibility for a task that it is effecting. Thus, adaptive automation is conceptually much more flexible in how it effects tasks, and dynamic in the frequency with which it can change status. These aspects of adaptive automation give it the potential to solve many problems that are created and/or not addressable with conventional automation, yet at the same time introduce a number of new human-machine interaction problems that must be empirically studied before this technology can be effectively applied to person-machine systems. The research described in the papers of this report describe contemporary research into a variety of issues in adaptive automation. These include: workload, situation awareness, cycles of automation, the issue of control and authority in adaptive automation, and alternative interface concepts for and adaptively automated crewstation. In addition, a set of prospective, theoretically derived principles and guidelines are presented.

Prospective Principles and Guidelines for the Design of Adaptively Automated Crewstations

Jeffrey G. Morrison, Ph.D.[1], David Cohen, and Jonathan P. Gluckman, Ph.D.
Naval Air Warfare Center - Aircraft Division, Code 6021, Warminster, PA 18974

ABSTRACT

Adaptive automation (AA) is a technology that has been proposed to tailor automation to human requirements. AA, when applied to the pilot-vehicle interface, is expected to minimize the negative effects of fixed automation while optimizing pilot performance. It is unclear, however, how AA should be designed to ensure optimal performance. This paper provides a review of an ongoing human factors research program that has the objective of providing a strong empirical foundation for the introduction of AA. A taxonomy for conceptualizing the design of AA systems is described and 13 prospective, research-based principles and guidelines for the implementation of AA are presented.

A prominent factor in the limiting of tactical aircraft performance is the inability of pilots to operate at the full potential of the aircraft (ref. 11, 12, 13, 16). One aspect of this problem is that pilots cannot process all the information presented to them in the limited time available. Adaptive automation (AA) technology has been proposed as a way to help pilots manage both information and task demands. AA is an approach to automation wherein the control of the onset, offset and form of automation in a person-machine system is mutually shared between the human and the machine. Thus AA differs from conventional (i.e., fixed or static) automation in two important ways: 1) AA can change the automation status of a task or function *autonomously*; and 2) AA may change the *functional characteristics* of tasks performed by the pilot.

The AA concept is based on the rationale that pilot performance may be optimized by managing the flow of information and task demands so the pilot's resources are appropriately allocated continuously over time. Specific task demands are selected and modified to ensure that the most critical tasks are attended to by the pilot and an optimal level of workload is maintained. The AA system would be sensitive to mission context – effectively adapting to both pilot and mission requirements (ref. 2, 6, 11, 12, 13, 22). To assess the validity of AA and explore potential pitfalls, the Navy has undertaken a basic research program to empirically assess the validity of this premise. The *Adaptive Function Allocation for Intelligent Cockpits* (AFAIC)[2] program is identifying critical aspects of AA and considering how these affect the perception and performance of pilot tasks. Areas investigated by the AFAIC program include: 1) the interaction of task demands in the cockpit and their effect on performance; 2) the contribution of simultaneously performed tasks to workload and situational awareness; 3) the impact of automation cycles on performance; 4) reliability of AA and pilot complacency; 5) the impact of operator versus computer control of automation; 6) interface structure and function for AA, and 7) training pilots to use AA (ref. 11, 12).

Figure 1 shows a taxonomy for how the AFAIC program has structured the impacts of AA on pilot functions.[3] This taxonomy enumerates AA methods based on three dimensions: the *philosophy* of automation invocation; the *strategy* of how pilot task demands should be adjusted; and the *stability* of the decisions being made by the pilot (ref. 6, 11, 12). AA can adopt a philosophy of either executing automation based on critical events occurring in the course of a mission or based on the real-time measurement of pilot state variables such as workload or performance (ref. 2). The automation can interact with the pilot in several ways. An entire task can be *allocated* between the pilot and the system. A task can be partitioned so that select aspects of the task are automated while others continue to be performed by the pilot. A final strategy is for a task to be *trans-*

---

[1]Technical correspondence should be addressed to the first author.

[2]AFAIC is a 6.2, Human Factors, Block-Funded program under the technical direction of NAWCAD, Warminster, PA, and the sponsorship of Mr. Jeff Grossman at NRaD, San Diego, CA.

[3]Gratitude to Mr. Edward Hitchcock for his help in refining this taxonomy.

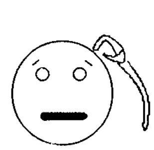
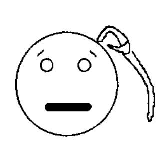
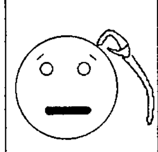
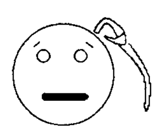
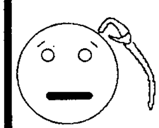
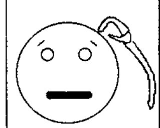
PHILOSOPHY:	*Critical Event Centered*			*Human Performance Centered*		
What factors cause changes in the automation status.	External events occuring in the environment causes automation status to change.			Real-time assessment of human performance causes automation status to change.		
STRATEGY:	*Allocation*	*Partition*	*Transform*	*Allocation*	*Partition*	*Transform*
How functions are changed by the automation.	Complete functions are shifted.	Parts of functions are shifted.	Cognitive demands of functions are changed.	Complete functions are shifted.	Parts of functions are shifted.	Cognitive demands of functions are changed.
DECISION STABILITY: *Stable* — Decisions require detection - minimal diagnosis.	+/?	-				
DECISION STABILITY: *Dynamic* — Decisions require diagnosis & strategy.	?/-	+/?				

Figure 1. Taxonomy for the implementation of adaptive automation.

formed so that the demands placed upon the pilot are changed while responsibility for the task remains with the pilot; (for example, a task's control requirement can be transformed from manual to voice and thereby fundamentally changing resource usage). The third dimension of this taxonomy, decision stability, reflects the inherent qualities of the decision-making components of a task after it is automated. The significant decision involved in a stable task is the detection of a relevant event. When an event is detected, the interpretation and action required is generally consistent and straightforward (e.g., a system monitoring task in which a button is pressed in response to a light or dial reading). Conversely, dynamic tasks have more complex detection rules and entail diagnosis and strategic decision-making (ref. 3). For example, a task where a change in a variable can mean different things at different times, and/or can require different control inputs would be dynamic. A fuel management task in a tactical aircraft, for which pumps are turned on and off to maintain a balanced fuel system, is a dynamic task.

The impact of AA is dependent on the specific combination of philosophy, strategy (ref. 11, 12) and decision stability of the function being automated (ref. 3, 5). Clearly the impact of AA on pilot performance is complex, and a significant amount of research is still required before this technology can reliably enhance pilot-vehicle performance. The symbols in Figure 1 reflect the data available on the impact of AA: a "+" indicates that there is data to suggest that the unique combination of dimensions is beneficial, a "-" indicates it is detrimental, a "?" indicates that there are mixed or conflicting results, and the "scratching head" symbol indicates that little or no known data are available. The remainder of this paper will briefly describe some of the AFAIC research as a way of starting to formulate design guidelines for the incorporation of AA into the crew station. The interested reader should refer to the technical reports for detailed discussions of these experiments.

Nature of Tasks. A prerequisite to the understanding of the behavioral impact of AA is an understanding

of how multiple tasks are performed concurrently. Resource theory and traditional single-versus-dual task research suggests that there is a fixed pool of resources that are allocated among the tasks. When all available resources have been allocated, any changes in task demands will be directly reflected by changes in performance (ref. 10, 15, 24). Early research in the AFAIC program showed that the modification of task demands significantly altered the way resources were allocated to individual tasks in a multi-task environment, and further that subjects change their performance strategies as a function of the relative changes in component tasks (ref. 11, 12, 13, 14). For instance, Morrison et al. (ref. 12, 13) found that increasing the driving frequencies of a continuous control task *improved* performance for a binary classification task with no significant change in tracking performance. This non-intuitive result suggested that subjects were shifting from a "frequency modulation" to an "amplitude modulation" tracking strategy. Further, it was suggested that the strategy used could be predicted by a behavioral reinforcement model. This interaction of task demands and performance suggests that a multi-task environment is best characterized as a single complex task with a *strategic* allocation of resources among the task components. Prospective guidelines derived from this research include:

1. Implicit pilot consent and minimal disruption of a pilots resource allocation strategy is best accomplished by the application of AA to more difficult/less reinforcing tasks in a task suite by virtue of pilots inclination to perform the more reinforcing or easier task.
2. To improve situational awareness for all tasks under the cognizance of the pilot, AA should be designed to equalize levels of task difficulty so that pilots interact equally with all tasks.

Cycles of Automation. Cycles of automation refers to the frequency with which automation is turned on/off over a period of time. There is a continuum of short to long cycles of AA, and what constitutes short or long cycles is dependent on the particular task being performed. Using variants of the NASA-Multi-Attribute Task (MAT) battery (Ref. 4) various AFAIC researchers have employed cycles of: manual control, automation of one of three tasks, and return to manual control. The tasks employed were: 1) system monitoring (a signal detection task characterized by stable decision making); 2) resource management (a cognitive-strategic task characterized by dynamic decision making); and a compensatory tracking task (a continuous manual control task characterized by stable decision making).

The frequency of automation transitions falls conventionally into two categories. In long-cycle adaptive automation, a function is automated for a long period of time, transfers to manual control for a short period, and then reverts to a long period of automation. In short-cycle adaptive automation, a function is cycled between manual control and system control more frequently, particularly if the technique used to determine transitions is susceptible to small changes in task demands or pilot workload. The AFAIC program operationally defined these terms in experiments to uncover the costs and benefits of each cycle type. The AFAIC program modified the standard MAT Battery so that it would be more suitable for adaptive automation experiments (see ref. 17 for specifics).

Parasuraman and colleagues (ref. 18) designated short-cycle automation as alternating 10 minute blocks of manual and automated control. Their study investigated the human performance impacts of such transitions in a multi-task environment by using a design in which subjects manually performed three of the tasks from the MAT battery simultaneously for the first block, followed by a block with manual performance on two of the tasks and the third automated (but monitored), and finished with a block of all tasks being performed manually again. Automation of the system monitoring or fuel management task led to improved performance in the tracking task and there was no evidence of a performance decrement for the return to manual control of the tasks following automation control (i.e., an automation deficit [ref. 25, 16]).

The capacity to monitor or "supervise" automation is an important indicator of system awareness as well as a source of workload. Subjects were instructed to monitor the automated tasks for deviations (each task was programmed with specific faults). The results of this study (as well as other AFAIC studies) revealed that the supervision of dynamic tasks was significantly worse than the supervision of more stable tasks. AA guidelines for short cycles of AA include:

3. Tasks involving continuous control (e.g., tracking) are most likely to exhibit performance benefits from

AA.

4. Use of relatively short cycles of AA has no negative consequences for manual task skills (i.e., an automation deficit).
5. The application of AA to tasks involving a vital diagnostic or history component will be susceptible to poor monitoring.

The issues arising from a consideration of both short-cycle and long-cycle AA are very similar: the effects associated with transitions and the ability to efficiently monitor a task during automation. Contrasts between the two cycle rates lie in the duration of the automation. Parasuraman et al. (ref. 21) examined long cycles AA by giving subjects 40 minutes of allocation of one of the three tasks, followed by 10 minutes of manual control of all the tasks and finally another 40 minutes of allocation. As the duration of AA increased, the detection of "non-optimal" AA decreased. On the return to allocation (after the inserted period of manual control), the detection of failures of the AA returned to a level comparable to that at the start of the experiment. Further, the progressive decline seen in the final period of automation occurred at essentially the same rate seen in the initial block of automation. Therefore, intermittent periods of manual control served to restore monitoring performance; minimizing the negative effects of extended automation. Supervision improved significantly for all three tasks, although there seemed to be a greater improvement for dynamic tasks. Therefore:

6. The performance benefits seen from allocation of tasks are transitory with extended periods of automation.
7. Use of intermittent periods of manual control during extended periods of task allocation will significantly improve the monitoring of the automation – restoring monitoring efficiency to the levels seen before allocation was initiated.
8. The periodic suspension of automation may be used for ensuring optimal performance, regardless of the type of task being automated. More substantial gains in monitoring performance can be expected for automating dynamic tasks.

Workload & Situational Awareness. NAWC has recently completed two studies that assess subjective workload and situation awareness (SA) (ref. 5, 3). The studies used a modified version of the MAT battery in which both allocation and partitioning strategies were implemented for fuel management and system monitoring tasks. The results of these studies showed the importance of decision-making stability in adaptive automation. Partitioning of a stable task (i.e., the system monitoring task) caused workload to increase (as measured by the NASA Task Load Index). Stability of the decision-making also affected the awareness of task performance during AA. Guidelines from these studies include:

9. Using a partitioning strategy with stable tasks will increase workload and lead to relatively poor awareness of automation performance.
10. There is a tradeoff between automation supervision and SA. The more dynamic a task, the less of a performance gain will occur when other stable tasks are automated; however, SA will be less impacted by automating stable tasks. The more stable a task, the greater the performance gain that will be realized by automating competing dynamic tasks; however, the greater the cost to overall SA.

Complacency and AA Reliability. Complacency, or the failure to adequately monitor an automated system, is a major concern when automating aircraft systems (ref. 25, 16). AA has been advocated as a means of minimizing complacency potential through the intermittent strategic adaptation of tasks between the human and machine components of the system. Parasuraman, Molloy & Singh (ref. 20) examined the development of complacency by manipulating instances of non-optimal automation within an automated system. The AA system had multiple levels of reliability – defined in terms of the percent of automation transitions that were not successful. The study used conditions in which a system monitoring task was automated with high, low or variable reliability. The results showed a significant development of complacency as measured by the failure to detect deviant automation, for the high and low reliability AA when compared to the variable reliability AA conditions. Thus, the changes in AA reliability succeeded in improving the monitoring of automation by the subjects. There were no performance differences for the tasks performed manually, suggesting that there were no differential costs associated with the monitoring of the automation. Therefore:

11. Transformation of a monitoring task by varying the rate of significant events may improve the efficiency of a supervisor, and therefore the performance of the overall system, relative to a system with constant, infrequent significant events. Changing the reliability of the automation may not generate significant performance effects on manually controlled tasks.

Training. Results of AFAIC investigations into training for AA (ref. 7, 12) suggest that the hybrid nature of AA would be best served by using a hybrid approach to training. Training should incorporate various feedback schemes depending on the specific combination of strategy and decision stability that will be employed.

12. The changes in performance requirements created by AA strategies dictate different kinds of feedback during training.

The AA Interface. The design of interfaces is central to the implementation of certain aspects of adaptive automation (e.g., transforming a task by changing the display format) and can have consequences on performance under automation control. Ballas, Heitmeyer, and Perez (ref.1) hypothesized that incorporating the elements of direct manipulation would enhance a subject's ability to monitor an automated task and therefore smooth the control transitions. Direct manipulation theory (ref. 9) posits that superior interfaces result from less information processing disparity between the user's intentions and the data provided by the machine.(i.e., distance) and more interaction with the application domain and the objects in it rather than through an intermediary (i.e., engagement). The results indicated that minimized distance and direct engagement mitigated some of the drawbacks associated with adaptive automation.

13. Generally, the direct manipulation interface lessened the impact of a transition to manual control when compared with interfaces characterized by greater cognitive complexity.

Future Directions. This paper has presented an overview of AFAIC adaptive automation research and its applicability to design guidelines. Clearly, there remains a great deal of work to be done before AA can be inserted into complex systems, such as those of the tactical aircraft cockpit, with some assurance that it will improve situational awareness and pilot performance. Examination of the current AA taxonomy illustrates that fundamental gaps in our knowledge exist; specifically concerning transformation strategies and the use of AA in human performance-based and hybrid critical event systems. Ongoing work in the AFAIC program will specifically focus on interface issues as these are likely to be critical to successful implementation of AA, particularly for transformation. Further, in order to meaningfully assess human performance based AA, it is necessary to first have means to measure real time performance, and this will be a major thrust of continued research at NAWCAD in the form of the *Automation Invocation Development* (AID) program.

References

1. Ballas, J. A., Heitmeyer, C. L., & Perez, M. A. (1992). *Direct manipulation and intermittent automation in advanced cockpits* (Final Report. NRL/FR/5534-92-9375). Washington, DC: Naval Research Laboratory.
2. Barnes, M. J. & Grossman, J. (1985). *The intelligent assistant concept for electronic warfare systems* (Final Report No. NWCTP5585). China Lake CA: Naval Weapons Center, December, 1985.
3. Carmody, M., & Gluckman, J. (1993). Task specific effects of automation and automation failure on performance, workload, and situation awareness. In R. S. Jensen (Ed.), *Proceedings of the Seventh International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University, The Department of Aviation, The Aviation Psychology Laboratory, (In press).
4. Comstock, R. J. & Arnegard, R. J. (1990). *Multi-Attribute Task Battery*. (Preliminary Technical Report). Hampton, VA: NASA-Langley Research Center.
5. Gluckman, J., Carmody, M., Morrison, J., Hitchcock, E., & Warm, J (1993). Effects of allocation and partitioning strategies of adaptive automation on task performance and perceived workload in aviation relevant tasks. In R. S. Jensen (Ed.), *Proceedings of the Seventh International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University, The Department of Aviation, The Aviation Psychology Laboratory, (In press).
6. Gluckman, J. P., Morrison, J. G., & Deaton, J. E. (1991). Complex task performance as a basis for

developing cognitive engineering guidelines in adaptive automation. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 116-120). San Francisco, CA: Human Factors Society.

7. Hilburn, B., Singh, I. L., Molloy, R., & Parasuraman, R. (1992). *Training and adaptive automation. I: Nonadaptive feedback training* (Technical Report No. CSL-N92-1). Washington, DC: The Catholic University of America, Cognitive Science Laboratory.
8. Hilburn, B., Wong, D., Molloy, R., & Parasuraman, R. (1992). *Training and adaptive automation. II: Amount of non-adaptive training* (Technical Report No. CSL-N92-3). Washington, DC: The Catholic University of America, Cognitive Science Laboratory.
9. Braune, R. J. and Trollip, S. R. (1982). Towards and internal model in pilot training. *Aviation, Space, and Environmental Medicine,* 53, 996-999.
10. Kahnemann, D. (1973). *Attention and Effort.* Englewood Cliffs,NJ: Prentice-Hall.
11. Morrison, J. G., Gluckman, J. P., & Deaton, J. E. (1991)[1]. *Program Plan for the Adaptive Function Allocation for Intelligent Cockpits (AFAIC) Program*, (Final Report No. NADC-91028-60). Warminster, PA: Naval Air Development Center.
12. Morrison, J. G., Gluckman, J. P., & Deaton, J. E. (1991)[2]. Human performance in complex task environments: A basis for the application of adaptive automation. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University, The Department of Aviation, The Aviation Psychology Laboratory, 96-101.
13. Morrison, J. G., Gluckman, J. P., & Deaton, J. E. (1992). *Human performance in complex task environments: A basis for the application of adaptive automation.* (Final Report No. NAWCADWAR-92032-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.
14. Morrison, J., Gluckman, J., Hitchcock, E., Barba, C., Wheery, R. & Glenn, F. (1993). *The effects of Adaptive Automation on performance in a complex, multi-modal task environment* (In preparation). Warminster, PA: Naval Air Warfare Center, Aircraft Division.
15. Navon, D. & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 63,* 81-97.
16. Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems* (Progress Report No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.
17. Parasuraman, R., Bahri, T., & Molloy, R. (1992). *Adaptive automation and human performance: I. Multi-task performance characteristics* (Final Report No. NAWCADWAR-92035-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.
18. Parasuraman, R., Bahri, T., Molloy, R., & Singh, I. (1992). *Adaptive automation and human performance: II. Effects of shifts in the level of automation on operator performance* (Final Report No. NAWCADWAR-92036-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.
19. Parasuraman, R., Hilburn, B., Molloy, R., & Singh, I. (1991). *Adaptive automation and human performance: III. Effects of practice on the benefits and costs of automation shifts* (Final Report No. NAWCADWAR-92037-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.
20. Parasuraman, R., Molloy, R., & Singh, I. L. (1991). *Performance consequences of automation-induced "complacency"* (Technical Report No. CSL-A-91-2). Washington, DC: The Catholic University of America, Cognitive Science Laboratory.
21. Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1992). *Training and adaptive automation. II: Adaptive manual training* (Technical Report No. CSL-N92-2). Washington, DC: The Catholic University of America, Cognitive Science Laboratory.
22. Rouse, W. B. (1986). Adaptive aiding for human/computer control. *Human Factors, 30,* 431-443.
23. Singh, I. L., Molloy, R., & Parasuraman, R. (1991). *Automation-induced "complacency": Development of a complacency-potential scale* (Technical Report No. CSL-A-91-1). Washington, DC: The Catholic University of America, Cognitive Science Laboratory.
24. Wickens, C. D. (1984). *Engineering psychology and human performance.* Columbus, OH: Merrill.
25. Wiener, E. L. (1988). Cockpit Automation. In: E. L. Wiener & D. C. Nagel (Eds.) *Human Factors in Aviation.* San Diego: Academic Press.

Effects of Allocation and Partitioning Strategies of Adaptive Automation on Task Performance and Perceived Workload in Aviation Relevant Tasks

J. P. Gluckman, M. A. Carmody, J. G. Morrison, E. M. Hitchcock
Naval Air Warfare Center - Aircraft Division - Warminster, PA

J. S. Warm
University of Cincinnati, Cincinnati, Ohio

ABSTRACT

Strategies for adaptive automation were studied in terms of their effects on performance and workload. The Multi-Attribute Task Battery (MAT) was employed for this study because it allows extensive inquiries regarding human information-processing in the presence of automation. Specifically, this study assessed the effects on performance and workload of automation strategies which vary the degree of operator control (no automation, partial or aided automation, and full automation) for tasks which differ on their stability with respect to time. A second objective of this experiment was to assess the use of a workload scale developed using aspects of the NASA-TLX and the SWORD techniques. The results suggest that the significant differences in operators ability to perform tasks, as well as their ratings of subjective workload, are affected by both the extent to which they maintain active control of the task and the stability of the task that is automated.

Adaptive automation (AA) has recently been proposed as an alternative approach to static, or traditional, automation. Static approaches typically dichotomize task control as either under full control of the operator (active or manual control) or fully automated. When a task is automated, traditional automation results in the operator being taken out of the control loop. This can lead to manual skills degradation, vigilance decrements, and loss of situation awareness (Parasuraman, Bahri, Deaton, Morrison & Barnes, 1992; Wiener, 1988). Unlike static automation, AA provides for automation status to be autonomously or semi-autonomously controlled by algorithms embedded in the automation system itself. Moreover, the form of the automation is more flexible than that of traditional approaches allowing various strategies of automated aiding to be employed: full automation (allocation), partial automation (aid), or task transformation. Algorithms governing both the onset and offset of adaptive automation and the automation strategy are responsive to a variety of factors. These may include operator initiated, as well as autonomous, system responses to real-time changes in operator-specific parameters (workload, performance, etc.) or external factors (task demands, system malfunctions, etc.). These factors have been proposed as the basis for an adaptive automation taxonomy (Morrison, Gluckman & Deaton, 1991[1,2]).

Benefits of AA are derived from the ability to keep operators in the control loop by altering levels of automation, and by tailoring the automation strategy as a function of the type of task being automated. A relevant theoretical distinction for this purpose is based upon the stability of the internal cognitive model that directs task decision making. This internal model is founded primarily in training and experience. It consists of patterns of associated events that direct a person's search for and interpretation of information (Braune & Trollip, 1982; Minsky, 1975). A <u>Stable</u> model refers to tasks in which the internal model, once learned, does not change across time. A <u>Dynamic</u> model, on the other hand, guides interaction with tasks where the significance of decision-relevant information does change over time. While both models rely on the operator's ability to detect and act, the degree to which they rely on a diagnosis phase of decision making is different. Tasks that use stable cognitive models require little diagnosis since the relevance of information does not change. Time-dependent tasks, which invoke dynamic cognitive models, rely heavily on the diagnosis phase since consequences and responses in the task change as a function of current/historical conditions. When applied to the AA problem, it is theorized that there will be consequences of systematically removing operators from direct control of stable versus dynamic tasks. Stable-model tasks should only be affected by task factors related to detection. Dynamic-model tasks, however, would be more severely affected by factors relating to strategy; particularly for allocation. This is because allocation would reduce opportunities for an operator to remain current with system changes and update his cognitive model (Carmody & Gluckman, 1993). For this class of tasks, an alternative automation strategy such as partitioning, in which the operator maintains some

level of task involvement, may lead to better performance because it would provide an increased opportunity for the operator to update his internal model.

One purpose of this study was to determine the performance effects of automation strategies on cognitive task type (i.e., stable versus dynamic-model tasks). Towards that end, an adaptation of the Multi-Attribute Task (MAT) Battery (Comstock and Arnegard, 1990) was employed. The MAT includes flight-relevant tasks that can be classified on the basis of their association with stable and dynamic internal cognitive models. The battery accesses three general information processing areas: perceptual-cognitive (a system monitoring task, **SM**), cognitive-strategic (a fuel management task, **FM**), and perceptual-motor (a tracking task, **T**) (Parasuraman, Bahri, and Molloy, 1991). The SM task requires subjects to monitor a panel of dials representing the temperature and pressure of two engines. When not automated, it is a stable-model task because the definition of signals and the responses to signals remains constant. The FM task, on the other hand, represents a dynamic-model task as the significance of changing fuel levels and determination of correct response were required under all conditions. In effect, the meaning of a pump being on could be positive or negative, depending upon the present state of supply tanks and status of other pumps regardless of automation status.

Instances of both allocation and aiding were generated for both the SM and FM tasks. When these tasks were allocated, a subset of responses was made by the AA. Tracking remained under manual control of the subject in all conditions. It should be noted, that aiding the SM task made it less stable, with regard to the subject's internal model. This is because aiding SM shifted decision making requirements such that both detection and diagnosis were necessary. Aiding SM required subjects to detect the occurrence of a signal **and** then determine whether they or the automation system should respond. Allocation of the SM task maintained a stable internal model because, as with manual performance, there was no change in the definition of signals and their responses. This dimension of the study enabled an examination of the possible role of automation design in manipulating the stability of the internal cognitive models.

It was predicted that when the SM task was allocated, performance on remaining tasks would be equal or greater to performance under conditions in which SM was aided, and the consequences of removing the subject from the more stable task would be minimal. With respect to the FM task, however, performance benefits were anticipated to be greater under the aided (partitioning) strategy, as this would preserve the subject's ability to remain current with the changes in task dynamics, thereby regularly updating the internal model.

The second major purpose of this study was to investigate the effects of alternative automation strategies on operators' perceived workload. Several popular techniques for measuring workload exist, but their application to attaining diagnostic information within the domain of adaptive automation and multi-task environments is quite limited. For example, the NASA-TLX is sensitive to overall workload changes as a function of changes in task demand (Gluckman, Becker, Warm, Dember & Hancock, 1990; Hart and Staveland, 1989). This scale also provides an evaluation of sources of overall workload by querying subjects on a variety of extrinsic and intrinsic workload dimensions. The extrinsic, or task related, factors include Mental Demand, Physical Demand, and Temporal Demand. The intrinsic factors, which measure the operator's affect as a function of interaction with a task include Own Performance, Effort and Frustration. However, the NASA-TLX does not provide information concerning individual contributions to workload of specific components in a multi-task environment or the relative changes in the source of workload under varying conditions of adaptive automation. An alternative measure of workload which could be used to assess these elements is the SWORD (Vidulich & Tsang, 1987). This technique allows the experimenter to structure direct comparisons between component tasks as well as between automated and non-automated phases. The SWORD provides only an overall evaluation of workload, however, and does not assess the contribution of component factors.

Given the diagnostic requirements of the AA problem, it was deemed appropriate to consider combining the attributes of the NASA-TLX and SWORD. In this case, the contributing sources of workload would be preserved from the NASA-TLX and merged with the pair-wise comparison procedure of the SWORD. The resulting SWORD-TLX rating scale was tested in the present study. In addition, a standard version of the NASA-TLX was administered to provide comparative workload data. It was expected that, in the case of the NASA-TLX, overall workload measures for both the SM and FM tasks would dichotomize on the basis of automation strategy. In the case of the SM task, it was expected that, while allocated SM would reduce workload, as compared to the manual control condition, aided SM would actually elevate workload. The latter was expected because of the added instability caused by the shift from stable to dynamic decision making. The application of both allocation and aided AA to the FM task, on the other hand, was expected to reduce

workload in comparison to the manual control condition. Further, allocation of FM would be perceived as less demanding than aided FM. Furthermore, with respect to the SWORD-TLX, while no specific predictions were generated, it was expected that the metric would differentiate between general changes such as automated state.

METHOD

Ten subjects (3 women and 7 men) volunteered from a pool of active duty military and Naval Academy cadets associated with the Naval Air Warfare Center, Aircraft Division, Warminster, PA. All subjects possessed little to no flight experience, had normal color vision, and 20/20 or corrected visual acuity. Subject ages ranged from 21 to 34 years, with a mean of 25.5 years.

A modified version of the MAT battery was used. The MAT was presented via a standard DOS 80386/-25Mhz computer equipped with a 19' VGA monitor. Changes in the standard MAT were made to produce allocation and partitioning automation strategies for both the SM and FM tasks. The changes included the addition of AA status boxes positioned in the lower right corners of each task window. The status box had the messages: "AUTO: FULL", "AUTO: AID" or "AUTO: OFF" printed in white for the allocation, aided (partitioned), or manual AA conditions. Fifteen seconds prior to the onset of a change in automation status for a task, a warning beep was sounded and the message for the new automation status was printed in yellow with brackets around the impending automation strategy. After the fifteen seconds elapsed, a second beep sounded and the appropriate status message was left in the task window. Under all conditions subjects performed the SM and FM tasks along with a first-order, compensatory, tracking task. This task required subjects to manipulate a joystick to keep a green circle centered over a target area. Tracking represented the main performance index used in this study. This task was the only one not manipulated, as past research has indicated that it is sensitive to changes in automated state (Morrison, et al., 1991[1,2]).

The SM task consisted of four vertically oriented scales representing temperature and oil pressure for two engines. Under normal conditions, a yellow pointer would fluctuate within set tolerances that were bounded by red lines and defined as the "normal" range. During manual operation of the SM task, the task was to monitor the four scales for an appropriate signal. A signal was defined as any time the yellow pointer on any of the scales moved completely out (either above or below) of the normal area. Upon detection of such a signal, the subject responded by depressing the appropriate response key for that scale. During full automation, all signals were responded to by the computer within 4 seconds. When the task was aided subjects were instructed that the computer would respond only to signals that went above normal range, they still had to respond to signals that went below normal. Across conditions, signals occurred at a rate of 12 per ten minutes with an inter-signal interval between 60 and 90 seconds. Both the scale in which a signal occurred and the direction of signals was random, with the restriction that half of the signals occurred in both the high and low directions. Reaction time to signals, as well as percent correct and false alarm data, were recorded.

The FM task consisted of two separate fuel systems linked only by two emergency transfer pumps. Each system contained a main tank, a reserve tank, and a supply tank. Directional fuel pumps, each with a set flow rate, connected the reserve and supply tank to the main tank, and the supply tank with the reserve tank. The status of each of the pumps was indicated by color-coded symbols (black = off, green = on, red = failed). Fuel level in each tank was graphically represented by green shading as well as an alphanumeric reading of fuel level in pounds presented below each tank. The subject's task was to maintain a specified range of fuel in each of the main tanks using the emergency pumps between the main fuel tanks only when there was no other way to maintain the desired level. Under manual operation of the FM task, subjects turned pumps on and off in response to fuel levels in the main tanks moving out of a predefined optimal range. The manner in which this was accomplished depended upon fuel consumption rate, active pumps, and which pumps had temporarily failed. Certain combinations of pump failures would also require the subject to use the emergency pumps to transfer fuel between the two main tanks. Under full automation (allocation), the system maintained the optimal fuel levels in the two main tanks by adopting the same strategy subjects had been briefed on in training. Under aided automation, the system operated in virtually the same manner as in allocated automation, with the exception the AA had no control over the emergency pumps. Under certain pump failures, the system would be incapable of maintaining the optimal fuel levels without the use of the emergency pumps, and subjects would need to intervene by operating the emergency pumps. The rational for this division of control was based on the expectation that emergency or non-standard situations would be those most critical for operator involvement, even during periods of automation. In all conditions, pump failures occurred at a rate of 12 per 10 minute period. These pump failures also occurred at random, with the restriction that half occurred

within each of the separate fuel systems. Root Mean Square Error (RMSE) of main tank fuel levels served as the performance measures in this task.

The experiment utilized a completely within-subjects design with 5 total conditions. These consisted of a control condition in which no automation was given for either the SM or FM task, and two task types (SM and FM) factorially combined with the two levels of automation strategy (full/allocation and aided/partitioning). Each trial consisted of three consecutive 10-minute periods. Except during the control condition, the onset of automation of the SM or FM task always occurred during the second period. The order of conditions was randomly assigned such that subjects could not anticipate the type of automation that would occur or the task to be automated.

All subjects were given a briefing package prior to the experiment and 45 minutes of training on the MAT. Subjects were run over a period of two days to avoid problems with fatigue. Two conditions were given the first day and three the second. Upon completion of each trial subjects were given a computerized version of the NASA-TLX, followed by a paper and pencil version of the SWORD-TLX. This scale consisted of a set of six comparison sheets, each corresponding to one of the workload contributors of the NASA-TLX (Mental Demand, Physical Demand, Temporal Demand, Effort, Own Performance, and Frustration). Subjects would then rate that contributor for the pair-wise comparisons of each task (Tracking, SM and FM), as in the SWORD procedure alone. A new set of comparisons was administered after each trial. For trials containing an automated period, subjects were asked to rate the relative demands of the tasks during the automated phase.

RESULTS

Performance: Tracking performance (RMSE) for all conditions over the three periods of each session is presented in Figure 1. Recall that tracking task performance served as the primary performance index for this study. As can be seen in the figure, tracking performance relative to the control condition was unchanged for both aided and full automation of the SM task relative to the manual condition. When the FM task was automated, however, tracking performance was better relative to the manual control, with the best performance occurring in the full automation or allocated condition. An analysis of variance confirmed these observations revealing only a significant interaction of automation condition by periods ($F = 4.64$; $p < .05$). Post-hoc Newman-Keuls tests revealed that full automation or allocation of the FM task was the only condition to generate a significant gain in tracking task performance. Performance under aided automation of the FM task was not significantly better than the other conditions.

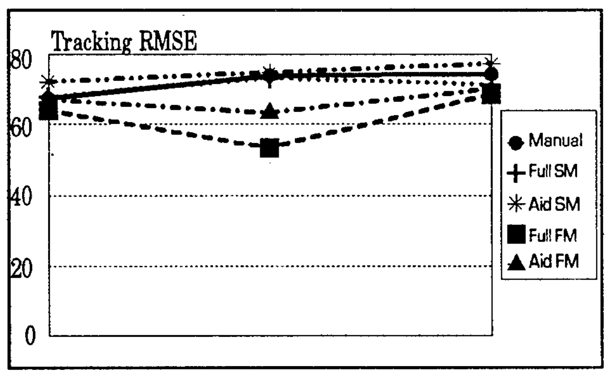


Figure 1. Tracking RMSE.

RMSE of the main tank levels of the FM task for periods one and three (pre- and post- automation) were also analyzed for all conditions and yielded no significant differences. Similarly, no significant differences were found for the reaction time data and the percent of correct detections for periods one and three of the SM task. Subject performance on the SM task was uniformly high, with low reaction times and few detection errors. These results indicate that across all conditions, subject performance on the FM and SM tasks were equal prior to the onset of automation and that no perseverative effects of automation existed.

Workload: Overall TLX values are shown in Figure 2. An analysis of variance of the data revealed a significant main effect for automation condition ($\underline{F} = 6.06$, $p < .001$). As can be seen in Figure 2, AA of both the SM and FM tasks resulted in automation strategy-specific effects on workload. Post hoc Newman-Keuls tests revealed that under full automation of the SM task, overall workload was reduced relative to the manual control condition. Moreover, when the SM task was aided, workload was significantly higher than all conditions except the control. With regard to the FM task, overall workload was lower for both full (allocated) and aided (partitioned) automation relative to the manual condition, with the lowest workload rating under the full automation (allocated) strategy. Post hoc tests revealed that the only significant workload difference for

automating the FM task occurred for the full automation of the FM relative to the aided SM task. With respect to the TLX subscales, no significant interactions between the TLX subscales and automation strategy or task were found. However, a main effect for subscales was found ($F = 13.45$, $p < .05$), and post hoc Newman-Keuls indicated that Mental Demand was rated significantly higher than all other subscales and Physical Demand and Frustration were rated as significantly lower than all other subscales but not different than each other. The remaining scales were not significantly different from each other.

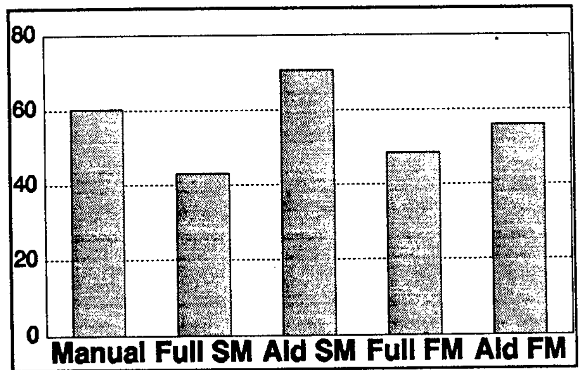


Figure 2. Overall Weighted TLX ratings.

SWORD-TLX: Significant main effects for automation strategy were found for each of the six workload factors. In each case, the pattern was quite similar, with both the manual control and the aided FM conditions rated as significantly less demanding than the others. The rated order of the remaining conditions (full FM, full SM and aided SM) varied. However, all were consistently higher than manual control and aided FM conditions, and not significantly different from each other. Further, significant interactions (automation by task) were found for three subscales: Physical Demand, Temporal Demand, and Effort ($F = 11.42$,$F = 2.56$,$F = 4.64$, $p < .05$). Figures 3 and 4 show the interactions for Temporal Demand and Effort respectively. As can be seen in these figures, trade-offs in task specific workload occurred as a result of the automation of tasks as well as the automation strategy used. In general, the three tasks (tracking, SM, and FM) were all rated as contributing equally under manual control. Moreover, when automation was used, workload was shifted away from the automated task.

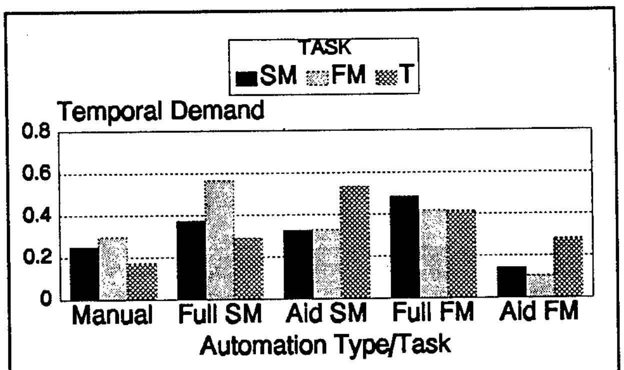


Figure 3. SWORD-TLX Temporal Demand.

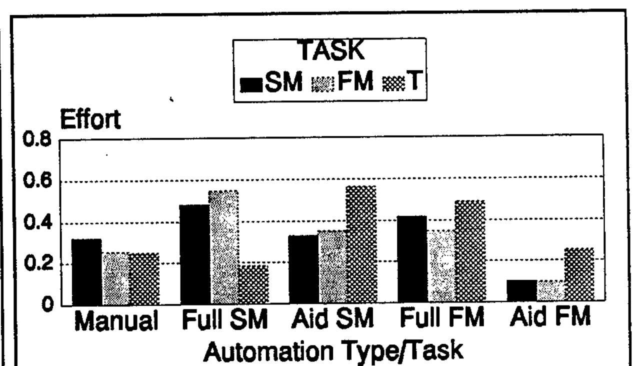


Figure 4. SWORD-TLX Effort Rating

DISCUSSION

The present study was conducted to evaluate 1) the effects of automation strategy on tasks which feature stable versus dynamic internal models, and 2) the utility of a new workload metric detecting changes in workload under conditions of AA within a complex multi-task battery. As predicted, tracking task performance was improved when the FM task was automated. However, a significant improvement only occurred during full automation (allocation). Moreover, no changes in tracking task performance were found when the SM task was automated. Taken alone, these results do not provide compelling evidence for the utility of the theoretical task distinction in question (cognitive model stability). These results, however, must be viewed in the context of several other factors. The NASA-TLX workload analysis indicated significant reductions in overall workload when the FM task was automated using both automation strategies (the greatest reduction found in the full automation condition). This result both confirms the benefit of automation and also indicates that the subjects experienced less workload when the dynamic FM task was partitioned. Similarly, as predicted by the model,

significant workload changes consistent with the decision-task distinction occurred when the SM task was automated. Full automation (allocation) resulted in a reduction of workload while partitioning of the SM task resulted in an increase in workload relative to no automation. The lack of supporting performance effects for the SM task is troublesome, but may be accounted for by the uniformly high subject performance in all conditions, suggesting a possible ceiling effect which negated the sensitivity of the tracking task performance in its ability to detect the AA effects. In this case, the SM task may have been too easy.

The SWORD-TLX results are quite interesting and indicate that the three tasks were perceived as contributing different types of demands, depending upon the automation strategy employed. As discussed above, specific trade-offs in the types of workload associated with each task occurred as a function of the automation strategy used. With future refinements to the SWORD-TLX, the results obtained in this study suggest that it may prove a worthy diagnostic tool for teasing out the exact sources of workload resulting from, or alleviated by, specific AA designs.

References

Braune, R. J. and Trollip, S. R. (1982). Towards an internal model in pilot training. *Aviation, Space, and Environmental Medicine, 53(10),* 996-999.

Carmody, M., & Gluckman, J. (1993). Task specific effects of automation and automation failure on performance, workload, and situation awareness. In R. S. Jensen (Ed.), *Proceedings of the Seventh International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University, The Department of Aviation, The Aviation Psychology Laboratory, (In press).

Comstock, R. J. & Arnegard, R. J. (1990). *Multi-Attribute Task Battery.* (Preliminary Technical Report). Hampton, VA: NASA-Langley Research Center.

Gluckman, J. P., Becker, A. B., Warm, J. S., Dember, W. N. & Hancock, P. A. (1990). *Subjective workload in dual-task monitoring*. Paper presented at the Southern Society for Philosophy and Psychology, Louisville, KY.

Hart, S. G. and Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshtaki (Eds.), *Human Mental Workload.* New York, NY: McGraw Hill. 211-277.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (ed.), *The Psychology of Computer Vision.* New York, NY: McGraw Hill, 211-277.

Morrison, J. G., Gluckman, J. P., & Deaton, J. E. (1991)[1]. *Program Plan for the Adaptive Function Allocation for Intelligent Cockpits (AFAIC) Program*, (Final Report No. NADC-91028-60). Warminster, PA: Naval Air Development Center.

Morrison, J. G., Gluckman, J. P., & Deaton, J. E. (1991)[2]. Human performance in complex task environments: A basis for the application of adaptive automation. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University, The Department of Aviation, The Aviation Psychology Laboratory, 96-101.

Parasuraman, R., Bahri, T., & Molloy, R. (1992). *Adaptive automation and human performance: I. Multi-task performance characteristics* (Final Report No. NAWCADWAR-92035-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

Vidulich, M. A. and Tsang, P. S. (1987). Absolute magnitude estimation and relative judgement approaches to subjective workload assessment. *Proceedings of the Human Factors Society 31st Annual Meeting,* 1057-1061.

Wiener, E. L. (1988). Cockpit Automation. In: E. L. Wiener & D. C. Nagel (Eds.) *Human Factors in Aviation.* San Diego: Academic Press.

Task Specific Effects of Automation and Automation Failure on Performance, Workload, and Situational Awareness

Meghan A. Carmody and Jonathan P. Gluckman
Naval Air Warfare Center, Aircraft Division, Warminster
Code 6021
Warminster, PA 18974-0591

ABSTRACT

The present study investigated the effects of automating different aviation-relevant tasks on human performance in regaining manual control following automation failure. The investigation employed a version of the Multi-Attribute Task (MAT) Battery which presents subjects three aviation-relevant tasks: a Compensatory Tracking task, a System Monitoring task, and a Fuel Management task. Specifically, this study examined the effects on performance, workload, and situational awareness of removing the human operator "from the loop" for long periods of time and then requiring him/her to suddenly reenter that "loop". Results indicated task-specific effects of automation on performance and situational awareness. Such effects are discussed with respect to the unique information-processing characteristics of the tasks involved, particularly the dynamic versus. stable nature of the internal cognitive model associated with decision-making within a task.

Modern aviation encompasses a complex realm of unique stimuli, and extraction of relevant information contained in these stimuli is the key to decision accuracy. Advances in cockpit automation have aided the pilot in this task, in part through workload reductions. However, the potential for automation-induced human error has raised concerns over possible losses in pilot situational awareness (Wiener, 1977; Wiener and Curry, 1980). Reducing pilot workload, while maintaining situational awareness, can only be accomplished by adopting a human-centered, as opposed to technology-centered, approach to cockpit automation. The present study was based on the premise that researchers cannot address this issue before understanding the unique attributes of human information processing within the semi-automated cockpit. The theoretical model of this process assumes the pilot has a variety of information sources regarding the state of the aircraft and the environment, received both directly and via an avionics system. The "decision process" of the automated component is argued to be guided by a program. Likewise, the decision process of the human operator is argued to be guided by an internal model. The internal model is a collection of learned patterns of events (Braune and Trollip, 1982; Minsky, 1975). These reduce information search by directing attention away from redundant/irrelevant cues. The model serves as the primary guide for the decision process, which consists of four stages: detection, diagnosis, decision, and execution (Flathers, Giffin, and Rockwell, 1982).

In detection, not all available information is attended by the pilot. The information to which the pilot attends is determined by the internal model guiding the visual scan. Upon detection of a fault, the pilot proceeds to diagnosis, searching for information to explain discrepancies. This is accomplished by examining plausible models for the situation, and re-adapting the scan to "test the fit" of such models. Ultimately, this may involve a model modification or transformation. The former involves adaptation of the current internal model to account for the new data; the latter involves selection of a more appropriate model to guide information sampling and decision making.(Barrett and Donnell, 1989). Once a diagnosis is made and a new model is operating, the pilot reaches, and then executes the decision (for detailed description of theoretical model see Carmody, 1993).

With respect to detection and diagnosis, this paper presents two hypothesized variations in decision processing by task. The first involves a task which is guided by a <u>Stable</u> internal model. This is one in which the information relevant to decision-making, particularly diagnosis, <u>does not</u> change across time. The second involves a task which is guided by a <u>Dynamic</u> internal model. This is one in which the information relevant to decision-making <u>does</u> change across time. The manner in which this task distinction operates in decision making, and why it is germane to task automation can be understood with respect to situational awareness.

Situational awareness (SA) has been defined by Endsley (1990) as "the perception of the elements in the environment within a volume of time and space [Level I], the comprehension of their meaning [Level II], and the projection of their status into the near future [Level III]". A certain degree of SA loss is expected to accompany long-cycled automation due to classic human vigilance problems, as well as those particularly noted with automated aviation tasks (Chambers and Nagel, 1985; Gluckman, Morrison, and Deaton, 1991; Parasuraman, 1987; Parasuraman, Bahri, and Molloy, 1991; Wiener and Curry, 1980; Wiener, 1988). The quality of loss, however, is argued to be related to the hypothesized stable and dynamic task distinctions. When the human operator is removed "from the loop" for long periods of time, return to manual control should differ with a stable versus dynamic model task, particularly if return-to-manual condition is sudden and/or unexpected, as in automation failure.

In the case of a stable model task, the weight of the decision process is upon the detection stage. Because the decision-relevant information in a stable model task does not change across time, removal of the human operator from that task results in loss of SA on a more perceptual level. Applying Endsley's (1990) definition, loss of SA in a stable model task should be most relevant in Level I, as the elements of Levels II and III, for the most part, do not change. Once detection occurs, the stable model can be called upon to guide the remainder of the decision process. With a dynamic model task, on the other hand, the weight of the decision process is upon diagnosis. If the human fails to monitor automation, SA loss is critical at deeper levels, as the decision-relevant information pertaining to those levels is changing. Therefore, when called to reenter the loop, the operator must not only detect discrepancies, but also **update** the internal model guiding diagnosis of the problem, as the established model may no longer be valid.

Two studies were conducted in order to examine aspects of performance and workload (Study I) and situational awareness (Study II). Study I employed the Multiattribute Task (MAT) Battery (Comstock and Arnegard, 1990). The MAT battery includes three aviation-relevant tasks which differ in cognitive type: a Tracking Task, a System Monitoring (SM) Task, and a Resource (Fuel) Management (RM) Task. This battery was selected for the present study because of the clear distinction between the System Monitoring and Resource (Fuel) Management Tasks in terms of stability of the associated internal models. The SM task required the subject to monitor a panel of four gauges, representing the temperature and pressure of two aircraft engines. When a signal (defined as any time a yellow pointer moved above or below the indicated normal range) was detected, the subject was to respond by depressing the appropriate key. This was a stable-model task, as the decision-relevant parameters defining signals and responses remained constant. In the RM task, on the other hand, the subject was to maintain a predetermined level of fuel in two tanks by turning on and off pumps (which periodically failed) to transfer fuel among a series of supplemental tanks. This was a dynamic-model task, as the task parameters and their defining relations changed across time. For example, whether turning pump 1 on was a positive or negative action depended on the current fuel and pump status. (For a more complete description of the MAT battery, see Comstock and Arnegard, 1990).

Additionally, effects due to Absolute (Abs) versus Comparative (Com) Judgement Types (Davies and Parasurman, 1982) were examined. This variable was added to manipulate the stability of the model within, as well as between, the tasks, in order to assure any effects found were due to internal model, rather than more general task differences. Comparative Judgement was established in both the SM and RM tasks by providing a visual referent (red lined) for the desired states. Absolute judgement provided only the defined limits (no visual referent). With Comparative Judgement in both the SM and RM tasks, the operator had a stable criterion in the visual referent. With the Absolute Judgement in both tasks, the operator had a memory-dependent criterion with potential for instability. Therefore, in the case of Comparative SM, one had a stable criterion within a stable model, whereas with Comparative RM, one had a stable criterion within a dynamic model. Furthermore, in the case of Absolute SM, one had a dynamic criterion within a stable model, while with Absolute RM, one had a dynamic criterion within a dynamic model.

Finally, Engagement Level was manipulated in the SM and RM tasks in order to test the idea that more task interaction is beneficial for model updates in a dynamic task. A High Engagement Level was defined as 12 signals/pump failures per 10 minutes. A Low Engagement Level was defined as 6 signals/pump failures per 10 minutes.

Given these theoretical constructs, it was predicted that manual operation of the RM task would interfere more with performance on the other tasks, due to the subject's need to maintain model updates. Likewise, automation failure of the RM task was expected to produce greater post-failure performance detriments in regaining manual control of RM, as compared to performance in regaining manual control of the SM task following automation of SM. This prediction was based on the premise that removing the subject from manual control of the RM task for long periods of time would lead to failure in updating the internal model. When suddenly required to regain manual control, the subject would at first be operating on an inaccurate model, adversely effecting performance.

Additionally, effects due to Judgement Type (Abs versus Com) were predicted to be task-dependent, with greater detriments seen when comparing Abs versus Com performance on SM relative to RM. It was expected that this variable would have more effect upon System Monitoring because it changed the quality , or fundamental character (in terms of stability) of that task. In the case of Resource Management, it only changed the quantity, or degree, of an already dynamic task.

Finally, it was expected that performance on the RM task would be enhanced under high RM engagement levels (because it would increase the model update opportunities), but not immediately following automation failure (because the increased occurence of events would add confusion in regaining Level II SA). With the SM task., on the other hand, it was predicted that SM performance would be enhanced, consistently, under high SM engagement levels, as the high engagement level would increase vigilance.

Study II employed a modified version of the Situational Awareness Global Assessment Technique (SAGAT) (Endsley, 1990). SAGAT collects objective situational awareness data for simulations. The procedure involves stopping the simulation at some random point in time, blanking the screen, and asking the subject a series of questions about the information present when simulation stopped. In the present investigation, subjects were queried on MAT task-specific questions on the basis of SA Level. The SM task was expected to show greater Level I (Perception) SA deficits following automation failure of the SM task. Furthermore, detriments in SA on the SM task were expected to be greater under Absolute Judgement Conditions.The RM task, on the other hand, was expected to show greater Level II (Meaning) SA deficits following automation failure of the RM task. It was not expected to be significantly effected by Judgement.

METHOD

Thirty-two volunteers from the Naval Air Warfare Center, Aircraft Division, Warminster served as subjects. All possessed little to no experience piloting aircraft, had 20/20 or corrected visual acuity, and normal color vision. Half the subjects served in the first study, the other half, in the second study. All subjects in Study I were male, with a mean age of 35. All but three of the subjects in Study II were male, and their mean age was 31.

Both studies employed the MAT battery (Comstock and Arnegard, 1990), presented via a standard DOS 80386 personal computer equipped with a 19" VGA monitor. Root Mean Square Error (RMSE) data was collected for the Tracking and Resource Management Tasks. Correct detections, false alarms, and reaction time data were collected for the System Monitoring Task.

Study I included a mixed between/within factorial design with 2 levels of the Between Subjects variable "Judgement Type" (Abs versus Com) by 3 levels of the Within Subjects variable "Automation" (Automated SM vs. Automated RM vs. Manual Control) by 2 levels of the Within Subjects variable "Engagement Level" (High vs. Low). Study II duplicated Experiment I in design, with the exception that "Engagement Level" was not manipulated (all tasks in all conditions were of High Engagement Level).

The procedure for Study I consisted of one training session, four experimental sessions, and three control sessions. Subjects were given a briefing package prior to their first training session. The package included detailed instructions on how to perform the subjective tests and the MAT battery. Subjects were interviewed on the first day to test their knowledge of the material. Each subject in Study I received 30 minutes of training and

three conditions on the first day, and a warm-up session and four conditions on the second day. All sessions lasted 30 minutes. During the first 20 minute period of experimental conditions, either SM or RM was automated. After 20 minutes into the session, automation failed (without warning). During the remaining 10 minute period, the subjects performed all three tasks manually. In control conditions, all three tasks were performed manually for the entire session. For data analysis, all sessions were divided into 6 continuous 5 minute blocks.

Procedure for Study II consisted of one training session, four experimental sessions (two in which SM was automated; two in which RM was automated), and two control sessions (full manual operation). As with Study I, subjects were given a briefing package prior to training. Each subject in Study II received 30 minutes of training and three conditions on the first day, and a warm-up and three conditions on the second day. Each subject was informed that at some point between 15 and 30 minutes into the simulation, the program would suddenly stop and the screen would blank, at which time the subject would be given the questionnaire.

RESULTS

Analysis of Variance was performed on all data, using Complete Statistical Software (Statsoft, 1992) Significant omnibus effects were subjected to post-hoc Newman-Keuls analyses. Percent correct data in both studies were subjected to Arcsine transformation.

Study I

Automation Effects:
Tracking performance during automation of the RM task was significantly better than during automation of the SM task ($F = 4.89$, $p < .05$). Manual tracking was not significantly different from either automated SM or automated RM conditions. There were no lingering (post-automation) effects on Tracking.

For the SM task, no meaningful significant effects were found for periods during or after automation. Analyses for the RM task, on the other hand, revealed significant improvements in RM performance during automation of the SM task ($F = 2.78$, $p < .05$). Furthermore, performance on the RM task deteriorated in the period following automation of the RM task. This was significant for block 6, when compared to the control condition ($F = 4.13$, $p < .05$).

Judgement Type Effects:
Detection performance in the SM task was significantly higher in both pre- ($F = 15.24$, $p < .05$) and post- ($F = 6.02$, $p < .05$) automation periods under Com versus Abs. Performance on the RM task was not significantly effected by this variable.

Engagement Level Effects:
Performance on the RM task in the control conditions improved significantly under High versus Low levels of RM. Detection performance in SM was only effected by this variable with respect to false alarm rate in block 5. False alarms were significantly greater ($F = 2.24$, $p < .05$) under High vs. Low SM levels.

Study II: Situational Awareness

There were no significant findings for Level I(Perception) SA questions about the RM task. However, with respect to Level I SA questions about the SM task, analysis revealed a significant main effect for condition ($F = 5.11$, $p < .05$) and a significant interaction between judgement type and condition ($F = 5.11$, $p < .05$). The post-hoc analysis indicated subjects had significantly better Level I SA for the SM task under Abs judgement. There was no significant difference between conditions in which RM was automated versus when SM was automated under the Com judgement. However, there was a significant difference between these conditions under Abs judgement, with Level I SA for the SM task superior following automation of RM. With respect to Level II

(Meaning) questions pertaining to the SM task, there was a significant interaction ($F = 4.65$, $p = .049$) between conditions in which the RM task versus the SM task was automated. Level II SA for the SM task was superior under the automated RM condition, but only under Absolute judgement. There was no difference between the conditions under Com judgement. Finally, Level II SA for the RM task was worse following automation of the RM task versus automation of the SM task or the control. This approached significance ($F = 4.18$, $p = .06$) when comparing automated RM with automated SM conditions.

DISCUSSION

The evidence from the present study offers strong support for general task differences with respect to automation. Furthermore, a close examination of the data indicates support for the theoretical task distinctions (on the basis of a stable vs. dynamic internal model), based on predicted outcomes.

As predicted, performance on Tracking was significantly better while the RM task was automated, supporting the idea that automation of manual control of the RM task, while preferable for maintaining SA, is more attention-demanding, due to continuous updating of changing model parameters. Furthermore, the Engagement Level findings for the RM task support predictions concerning RM benefits from increased subject involvement with the task resulting in more frequent model updates and a more accurate model. Additionally, while automation of a both tasks resulted in a post-automation drop in performance in regaining manual control of the once-automated tasks, these effects were only significant in the case of the RM task. Also as predicted, only performance on the SM task was significantly effected by the manipulation of the internal model within the task (Abs vs Com). Examined together, these two latter findings support the prediction that the less stable the model guiding a task, the more detriments in performance it will sustain immediately following a long period of automation.

The findings from Study II augment those of Study I, particularly with respect to Level II (Meaning) SA. As predicted, the less stable the model guiding a task, the greater the detriments to Level II SA during automation of that task. Without an understanding of the meaning of changing parameters within a dynamic-model task, one has little hope for an accurate model upon which to base decision-making performance when regaining manual control following automation.

In summary, the data provides strong support for further examination of aviation task distinction on the basis of the stability of the internal cognitive model, and how this effects performance within a semi automated cockpit. Although the data are not entirely conclusive, the study provides a firm foundation upon which to build a line of research. Future research along this vein will concentrate upon refining the sensitivity and fidelity of the measures, as well as the representation of subjects.

REFERENCES

Barrett, C.L. and Donnell, M.L. (1990). Real time expert advisory systems: Considerations and imperatives. *Information and Decision Technologies*, 16, 15-25.

Braune, R.J. and Trollip, S.R. (1982). Towards an internal model in pilot training. *Aviation, Space, and Environmental Medicine*, 53(10), 996-999.

Carmody, M.A. (1993). *An Examination of Theoretically Distinct Internal Cognitive Models and Their Role in Human Performance within a Semi-Automated Cockpit.* NAWCADWAR Tech Doc (in progress).

Chambers, A.B. and Nagel, D.C. (1985). Pilots of the future: Human or computer? *Communications of the Association for Computing Machinery*, 28, 1187-1199.

Comstock, J.R. and Arnegard, R.J. (1990). Multi-attribute task battery. NASA Langley Research Center, Hampton, VA. Report in progress.

Davies, D.R. and Parasuraman, R. (1982). *The Psychology of Vigilance*. London: Academic Press.

Endsley, M. R. (1990). *Situation Awareness Global Assessment Technique (SAGAT):Air to Air Tactical Version: User's Guide*. Northrop Corporation, Hawthorne, CA. NOR DOC no. 89-58 REV A.

Flathers, G.W., Giffin, W.C., and Rockwell, T.H. (1982). A study of decision-making behavior of aircraft pilots deviating from a planned flight. *Aviation Space and Environmental Medicine*, 53(10), 958-963.

Gluckman, J.P., Morrison, J.G. and Deaton, J.E. (1991). Complex task performance as a basis for developing cognitive engineering guidelines in adaptive automation. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 116-120). San Francisco, CA: Human Factors Society.

Hart, S.G. and Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshtaki (Eds), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston(ed.), *The Psychology of Computer Vision* (pp 211-277). New York, NY: McGraw Hill.

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors*, 29(6),695-706.

Parasuraman, R., Bahri, T., and Molloy, R. (1991). Adaptive automation and human performance: I. Multi-task performance characteristics. (Technical Report No. CSL-N91-1), Cognitive Science Laboratory, The Catholic University of America, Washington, D.C.

Statsoft (1992). Complete Statistical System (CSS): STATISTICA. Tulsa, OK.

Weiner, E.L. (1988). Cockpit automation. In Weiner, E.L. and Nagel, D.C (Eds), *Human Factors In Aviation*. San Diego, CA: Academic Press, Inc.Wiener, 1977; Wiener and Curry, 1980; Wiener, 1988).

Weiner, E.L. and Curry, R.E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23(10), 995-1011.

EFFECT OF TASK LOAD AND TASK LOAD INCREMENT ON PERFORMANCE AND WORKLOAD

P.A. Hancock and G. Williams

Human Factors Research Laboratory
164 Norris Hall
172 Pillsbury Drive, S.E.
University of Minnesota
Minneapolis, MN 55455

INTRODUCTION

The goal of adaptive automated task allocation is the 'seamless' transfer of work demand between human and machine. In essence, this is a replication of the strategy which allows consciousness to perform and integrate multiple tasks while retaining a sensation of perceived unity (Minsky 1984). Clearly, at the present time, we are far from this objective. Current systems, particularly high-performance, single-seat aircraft demand continual attention switching and display 'scanning' in order to maintain an adequate awareness of situation and required action. One of the barriers to achieving effortless human-machine symbiosis is an inadequate understanding of the way in which operators themselves seek to re-allocate demand among their own personal 'resources.' We have begun to address such issues through an examination of workload response, which scales an individual's reaction to common levels of experienced external demand, e.g., take-off, nap-of-the-earth, carrier landing, etc. Despite a considerable history of workload research (e.g., Gopher & Donchin, 1986; Hancock & Meshkati, 1988; Moray, 1979) there is much that remains uncertain about mental workload and the way in which workload characteristics affect performance strategy.

Understanding workload response is an importanı facet of development in adaptive automated task allocation, since a key contemporary question is the invocation and extraction procedures through which operational mode changes are made. This enquiry is part of a general examination of cues, both environmental and operator-based, upon which the transfer between manual and automated performance is achieved. At present, of course, this assumes discrete distribution of task demand between systems and operators and does not embrace the complementarity notion of task allocation as articulated by Jordan (1963). The way in which such sharing or 'partitioning' of task can be achieved is considered in further experiments in the present series. (see also Parasuraman, Bahri, Deaton, Morrison, and Barnes, 1990). It is anticipated that the invocation and extraction process will be a strong determinant of the acceptance or rejection by pilots of re-configurable interface structures for adaptive task allocation.

PROPERTIES OF THE WORKLOAD RESPONSE

The present experiment is predicated upon some earlier observations by Hancock and Chignell (1987) on workload response as a key facet of adaptive allocation. In that work, we identified a number of characteristics of workload that could be considered in a manner similar to a response following on a varying level of task demand. represented as an analog signal. These characteristics are illustrated in Figure 1. Workload level is plotted against time and provides a number of discrete regions and trends. The regions are labeled overload and underload, with a blank region of acceptable load between. We assume here, as has become *a leitmotif* for all of adaptive allocation, that prolonged residence in règions of maladaptive load (either underload or overload) is detrimental to operator performance and will result in an increase in error and decrease in performance speed In the present work, the concern is with the mitigation of overload, although in principle results can be used for the amelioration of underload also. What is identified by the hashed regions in Figure 1 is the summation of time and intensity spent in unacceptable regions. In the past it has been assumed that a workload 'redline' exists which cannot be fractured at any cost. However, there are many situations in flight (especially combat) which put pilots *in extremis* and cannot simply be 'avoided.' Therefore, consideration of the total time and level of workload spent 'out of the workload envelope' is a critical consideration and one that has only rarely been explored (Hancock, 1989).

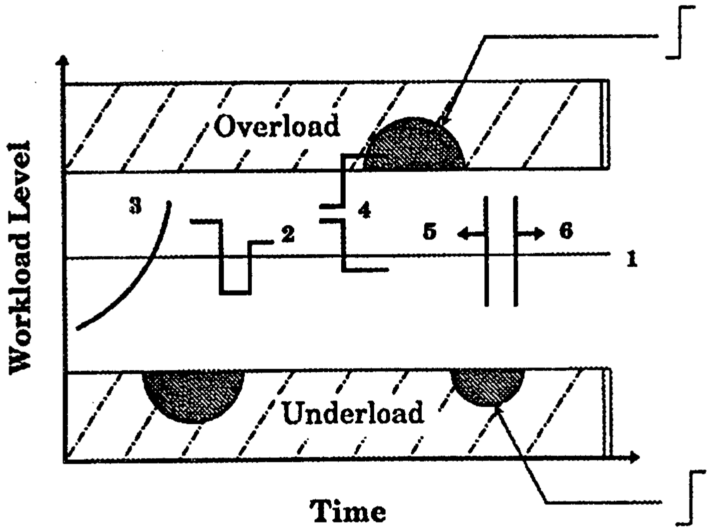


Figure 1. Facets of workload response as a function of variation in task demand.

Within the region of acceptable performance we have traditionally been concerned with the absolute value of workload as shown by level (1) above. However, as workload has been shown to dissociate with performance, and workload evaluation in highly demanding or highly boring conditions is difficult to assess, absolute level of workload is not always the most useful metric. However, there are several other facets of task demand workload relationship that may be of use especially in relation to adaptive allocation. Some of these linkages have already been shown to be influential, such as past history, see (5) above (Matthews, 1986; Miyake, Hancock, & Manning, 1992). These findings suggest that other facets such as future expectation (6) (see Harris, Hancock & Arthur, 1993) and level and location of recovery (4) may also prove of value. In the present experiment we examine the facet of workload here labeled (2), that is level of workload combined with increment in workload. In actuality, this is also a manipulation of the trend illustrated as (3) being rate of change of workload. As humans are frequently more sensitive to change and rate of change rather than absolute level, we have a rationale for belief that increment of workload is an important variable in influencing performance response. The present experiment is also part of a general programmatic investigation that we have pursued on workload transition events (see also Miyake, Hancock, & Manning, 1992). Recognition of the importance of workload transitions is clearly growing (see Howell, 1992; Huey & Wickens, 1992; Warm, 1992). However, as yet relatively few experimental findings examine the assumption that manipulations of task loading are followed by concomitant change in workload, and it is change in the level of such workload that promises to be a key facet in initiation and cessation of automation in high demand environments.

METHOD

Experimental Participants

The participants in the present experiment were fifteen student volunteers from the University of Minnesota. There were nine males and six females. The mean age of the sample was twenty-four with a standard deviation of five. Subjects were volunteers and all were in professed good health at the time of testing.

Experimental Procedure

The experimental platform to test the present hypotheses was MINUTES (MINnesota Universal Task Evaluation System). This facility was developed at the University of Minnesota's Human Factors Research Laboratory and is freely available (see Harris, Hancock, Arthur & Manning, 1992). This environment consists of three major subtasks namely; monitoring, resource management, and tracking, each of which can be controlled through precoded scripts. Subjective assessment of workload was collected by having subjects complete SWAT (Subjective Work Assessment Technique) tests which appeared in a window within the MINUTES environment. Subjects were provided with training to become familiar with the tasks that would be completed during the experimental sessions. The tasks were completed using a joystick and keyboard controls. The monitoring task required keyboard responses to indicator lights and gauges. Resource management required monitoring and control of fuel tanks and pumps to maintain constant target level of fuel in two of the five tanks. Tracking required joystick manipulation to maintain a crosshair at the center of a display. Task load baseline and increment levels were determined by varying frequency of light and gauge state changes, frequency and duration of pump failures, and changes in the gain of tracking and sensitivity of joystick as detailed below. The SWAT tests provided subjective assessment of time load, stress and mental effort. An illustration of the MINUTES interface is shown in Harris, Hancock and Arthur (this volume, figure 1).

The experimental protocol employed a within subject design. The experiment itself consisted of completing three sessions each lasting twelve minutes. Each session consisted of three sub-routines which were made up from a baseline level of task load lasting 100 seconds, followed by the three SWAT scales. The same baseline level task plus an incremental load (100 seconds) was then presented, also followed by the SWAT tests. The baseline level and incremental levels are explained below.

i) Baseline Level of Task Demand Conditions Three baseline rates were used for each of the components of the MINUTES task. The event rates are illustrated in Table 1. Each event refers to an entry in the script that creates change in state to either the monitoring task or the resource management task. Eight tracking levels were used for controlling the demand of the tracking task.

	Baseline	Baseline plus incremental		
		Low	Medium	High
Low	2 - 1 - 1	6 - 3 - 3	9 - 4 - 5	12 - 5 - 6
Medium	5 - 2 - 2	9 - 4 - 5	12 - 5 - 6	15 - 6 - 7
High	8 - 3 - 4	12 - 5 - 6	15 - 6 - 7	18 - 7 - 8

Table 1. Task Combinations
(a-b-c : a - monitoring events; b - resource management events; c - tracking value)

ii) Incremental Demand Conditions Three incremental levels were combined with the baseline task levels to form nine experimental conditions (Table 1). The rationale behind the size of the increment was to provide task levels in order to compare relative increase in load with absolute task level (e.g. low baseline level plus medium incremental is equivalent in difficulty to medium baseline level plus low increment.). Equivalent task loads are indicated by identical shading in Table 1.

RESULTS

Data were analyzed individually for each component task i.e. monitoring, resource management, and tracking. Each set was analyzed using a 3 baseline level (low medium and high) by 3 incremental level (low, medium and high) by 2 before/after increment (before (i.e. a baseline level) and after (i.e. a baseline plus an increment)) ANOVA with repeated measures.

i) Tracking Tracking performance is reported in RMS error units. Rather than compare the tracking data for the entire 100 second period it was decided that the most pertinent approach for this experiment was to consider tracking data from the last 20 seconds of the baseline task conditions and from the first 20 seconds of the baseline condition plus an increment. One main effect was found in the tracking data for the before and after condition, $F(1,8)=13.047$, $p<0.01$. The mean for the before condition was 19565.06, and for the after condition 32200.03. An interaction between increment levels and the before/after conditions produced significant effects $F(2,16)=5.248$, $p<0.05$. The interaction between baseline level and before/after condition was significant, $F(2,16)=7.395$, $p<0.005$. Post hoc t-tests were carried out on the before/after data for the baseline level conditions. These tests proved significant for the high baseline condition $t(8)=2.714$, $p<0.05$, and for the medium baseline condition $t(8)=2.898$, $p<0.05$.

ii) Monitoring Three sets of data were collected from the monitoring data: response time, response omission, and false alarms. The response time data were based on the response time for correct responses in seconds. For response time there was a significant main effect in the before versus after conditions, $F(1,14)=6.510$, $p<0.05$. The mean response time for the before condition (i.e. for a baseline) was 1.30 seconds and after condition (i.e. baseline plus an increment) the mean time was 1.42 seconds. The same significant main effect was evident in the false alarm data $F(1,13)=8.199$, $p<0.05$. The mean number of false alarms for the baseline conditions was 0.53 and for the baseline plus incremental condition was 1.03. This suggests that an increase in the task load produces an increase in both the time to react correctly to a monitoring cue and also the number of false responses both of which reflect deterioration of capability.

The data that produced the most interesting result were from the signal response omissions. The misses were converted to proportion scores representing the total number of misses with respect to the total number of possible correct responses. Two main effects proved significant: The before and after conditions $F(1,14)=8.617$, $p<0.05$ with means of 0.31 and 0.37 respectively; and the level of baseline task load conditions, $F(2,28)=32.801$, $p<0.00$ with means of 0.21, 0.4, 0.41 for low, medium and high baseline respectively. Two interactions also proved significant. The first was for the level of baseline rate and before/after increment $F(2,28)=4.106$, $p<0.05$, and the second was for the increment level and before/after condition, $F(2,28)=4.036$, $p<0.05$. Post hoc t-tests on the first interaction (i.e. the differences between the baseline values and baseline values plus increment for each baseline level) produced significant results for the differences between the low baseline condition and both the medium baseline condition $t(14)=1.995$, $p<0.05$, and the high baseline condition $t(14)=2.926$, $p<0.05$.

iii) Resource Management No significant results were found within the resource management data under any of the conditions.

iv) SWAT The three sets of SWAT data were analyzed: Time load, Stress and Mental effort. The SWAT data produced two significant main effects for each of the SWAT tests. This was for the before/after condition and for the level conditions. The means for each of SWAT response showed a trend of increasing with respect to work load for each of the scales. The results of each ANOVA are presented in Table 2.

Main Effect	Time Load			Mental Effort			Stress		
	F:	df	*p:*	F:	df	*p:*	F:	df	*p:*
Before/ After	8.576	1,13	<0.05	47.326	1,13	<0.001	19.519	1,13	<0.001
Baseline level	11.308	2,26	<0.001	11.308	2,26	<0.001	5.260	2,26	<0.05

Table 2. Significant main effects for the subjective workload responses.

Post hoc t-tests for each of the SWAT tests were applied and the significant results of these are shown in Table 3.

Level	Time Load			Mental Effort			Stress		
	t:	df	*p:*	*t:*	df	*p:*	*t:*	df	*p:*
Low vs Medium	3.606	13	<0.005	1.710	13	ns	1,422	13	ns
Low vs High	4.315	13	<0.001	4.678	13	<0.001	2.797	13	<0.05
Medium vs High	1.935	13	ns	2.156	13	<0.05	2.120	13	ns

Table 3. Results of the post hoc t-tests for the baseline effect on the subjective workload subscales.

DISCUSSION

The overall tenor of the present results indicate that the primary driver of performance is the absolute level of task demand over the increment in that demand. However, we must temper this observation because of the number of significant interactions observed. For example, in the tracking data, there was a significant modification of the before versus after increment effect because of the baseline level. It is critical to note, however, that the original baseline levels, i.e., the before conditions, do not exhibit a simple increase in RMS error with baseline demand. The interaction effect consequently seems to represent a threshold characteristic where it is the combination of an increment over a high baseline that triggers a non-proportional increase in RMS error. This is further clarified by the before/after, increment interaction. Here we see a differential increment effect which initially might lead us to support a case for the influence of such a manipulation. However, examination of the pre-increment baselines indicates that under the high increment condition the baseline was depressed such that the interaction appears. This suppression of baseline militates against a strong support for an increment effect here in tracking.

Further support for the task demand level primacy is seen in the monitoring data. The only significant effects in response time and false alarms reflect this demand characteristic. The pattern for signal omission is somewhat more complex. While the before/after pattern is maintained, an interaction occurs because of the effects in the low baseline condition. The difference between the before versus after comparison are exacerbated in the low baseline condition because of the low frequency of misses in the before condition. Again, as with tracking, we favor an explanation that revolves around a suppression of baseline effect rather than emphasizing the increment effect, since the latter influence did not percolate through all baseline levels. Also, the tracking suppression occurred at high baseline levels compared with the suppression in monitoring signal omissions at the low baseline level. This inconsistency argues against strong support for incremental influences.

Our conclusion is further buttressed by analysis of the workload data. Each of the SWAT subscales ubiquitously showed the before versus after difference and main effects for baseline load were evident in all scales. Also while all pairwise comparisons of workload response under baseline manipulations did not reach significance, the low versus high baseline conditions were always reliably distinguished. Overall our result confirm the primacy of absolute task load over incremental effects.

SUMMARY AND CONCLUSIONS

Our introduction posed the question of what characteristics of task demand that workload might be most sensitive to. In extension we suggested that outcome characteristics of workload (illustrated in Figure 1) can relate directly to the question of mode control transfer between automation and manual activity. We proposed to drive workload via manipulation of task demand level and increment of that level. Of the two, the first seems distinctively more potent. Of course some caution is necessary. First, it is possible that present levels of difficulty and increments on that difficulty were not sufficiently differentiated to elicit effects. In essence, the sensitivity of the measure argument will always be with us (Poulton, 1965). In addition the concern of our overall research program focuses on the cues upon which automation is initiated. Hence, additional work is already being performed in a

multi-task simulation facility in which the level of subtask difficulties have been magnified and automation invocation strategy observed. Such results will serve to establish the reliability of the information reported here.

ACKNOWLEDGMENTS

This research was supported by the Naval Air Warfare Center through N/N62269-92-C-0211 and NASA Ames Research Center through Grant NAG 2-749. Jeff Morrison and John Gluckman were the NAWC contact monitors and Sandra Hart was the NASA Grant Technical Monitor. The views expressed are those of the authors and do not necessarily represent those of the named agencies. We thank Erik Arthur for making alterations to the MINUTES program for the purpose of this experiment and Sue Chrysler for her assistance in all general areas.

REFERENCES

Gopher, D., & Dochin, E. (1986). Workload: An examination of the concept. In: K. Boff, L. Kaufman, and J.P. Thomas, (Eds). *Handbook of perception and human performance.* (pp.41:1-49). New York: Wiley.

Hancock, P.A. (1989). The effect of performance failure and task demand on the perception of mental workload. *Applied Ergonomics*, **20**, 197-205.

Hancock, P.A., & Meshkati, N. (Eds). (1988). *Human mental workload.* Amsterdam: North-Holland.

Harris, W.C., Hancock, P.A., Arthur, E., & Manning, C. (1992). *Minnesota universal task evaluation system* [MINUTES]. Technical Report 92-02, NADC, Warminster, PA.

Harris, W.C., Hancock, P.A., & Arthur, E. (1993). *The effort of taskload projection on automation use, performance, and workload.* Proceedings of the Seventh International Symposium on Aviation Psychology, Columbus, OH, April.

Howell, W.C. (1992). Engineering psychology in a changing world. *Annual Review of Psychology*, **44**, 231-263.

Huey, B.H., & Wickens, C.D. (1992). (Eds.). *Workload transition: Implications for individual and team performance.* National Academy Press: Washington, D.C

Jordan, N. (1963). Allocation of functions between man and machines in automated systems. *Journal of Applied Psychology*, **47**, 161-165.

Matthews, M. (1986). The influence of visual workload history on visual performance. *Human Factors*, **28**, 623-632.

Minsky, M. (1984). *The society of mind.* New York: Simon and Schuster.

Miyake, S., Hancock, P.A., & Manning, C.M. (1992). *Effects of prior task load level on subsequent workload and performance.* Paper presented at the 36th Annual Meeting of the Human Factors Society, Atlanta, GA, October.

Moray, N. (1979). (Ed.). *Mental workload: Its theory and measurement.* New York: Plenum Press.

Poulton, E.C. (1965). On increasing the sensitivity of measures of performance. *Ergonomics*, **8**, 69-76.

Warm, J.S. (1993). Vigilance and target detection. In: B.H. Huey and C.D. Wickens. *Workload transition: Implications for individual and team performance.* National Academy Press: Washington, D.C

THE EFFECT OF TASKLOAD PROJECTION ON AUTOMATION USE, PERFORMANCE, AND WORKLOAD

W. C. Harris, P. A. Hancock, and E.J. Arthur

Human Factors Research Laboratory
164 Norris Hall
172 Pillsbury Drive, S.E.
University of Minnesota
Minneapolis, MN 55455

INTRODUCTION

Just because you can automate a task doesn't mean you should. In conditions where system performance is not demonstrably better during either automatic or manual modes, the question of when and how control is transferred is an important operational issue (Hancock, 1992; Weiner & Curry, 1980). Loss of situation awareness and the perceived loss of control during automation suggest that manual control is preferable under some circumstances. However, automation is seen to be of particular value during periods of high taskload that threaten to exceed operator capacity (Billings, 1991, Weiner & Nagel, 1988). Operator control over automation initiation keeps the pilot in the loop, but as a result, pilots might become less likely to engage automation at the very time it is projected to be used (Parasuraman, Bahri, Deaton, Morrison & Barnes, 1990). The ability to switch between manual and automatic control produces performance superior to either static automation or manual control by itself (Harris, Hancock, Arthur & Caird, 1991). Thus although the necessity to switch between automation and manual control adds an additional task when performance is deteriorating, optional automatic can then improve performance.

The greatest challenge for automation in these circumstances is the transition between manual and automatic control modes. Performance deterioration has been observed during rapid, unexpected workload increases (Hancock & Williams, 1993). Determination of the conditions that minimize performance deterioration during the transition between manual control and automation would appear to be an important consideration when deciding the situations where adaptive automation is likely to be most beneficial. The present experiment compares multi-task, optional automation performance during periods when upcoming taskload information is available with performance during periods when taskload projections are not available and examines the effect of fatigue on operator's use of optional automation.

METHOD

Participants

Eight right handed University of Minnesota students participated in the study. They were introductory psychology students and received course credit for participation. All were in professed good health at the time of testing.

The Minnesota Universal Task Evaluation System (MINUTES)

The performance task used in this study was the Minnesota Universal Task Evaluation System (Harris, Hancock, Arthur, & Manning, 1992), a revised version of the Multi-Attribute Task Battery (Comstock & Arnegard, 1992). Each of these are multi-task, generic representation of complex systems. The MINUTES interface illustrated in Figure 1. Among the multi-task elements, monitoring tasks included a response when a green light extinguished, a response when a light to the right of the green light turned red, striking one of four keys that indicated when the corresponding gauge in the monitoring panel had moved beyond one "hash mark" from the center, and "diagnosing an engine problem". The "engine problem" began with the onset of the yellow "master warning" light and two gauges moving "out of range". Four combinations of gauge deviations were presented. The participants task was to strike to extinguish the master caution light and to then strike one of four engine problem keys to identify the gauge deviation pattern. The MINUTES also included a resource management task (lower center portion of the screen) which required participants to activate or inactivate the pumps connecting the tanks, which are represented by the small

squares in the screen representation. The goal was to maintain the fluid levels in tanks A and B as close to 2500 gallons as possible. The rates of flow from tanks A and B are larger than pumps 2 and 4 can provide, thus participants must develop a strategy that involves the periodic use of tanks C and D. Task difficulty can be increased by introducing pump failures which are indicated when the squares representing pumps turn red. In the tracking task, participants used a joystick to maintain the circle within a box in the center of the screen. The tracking gain was set at 45 during all experiments.

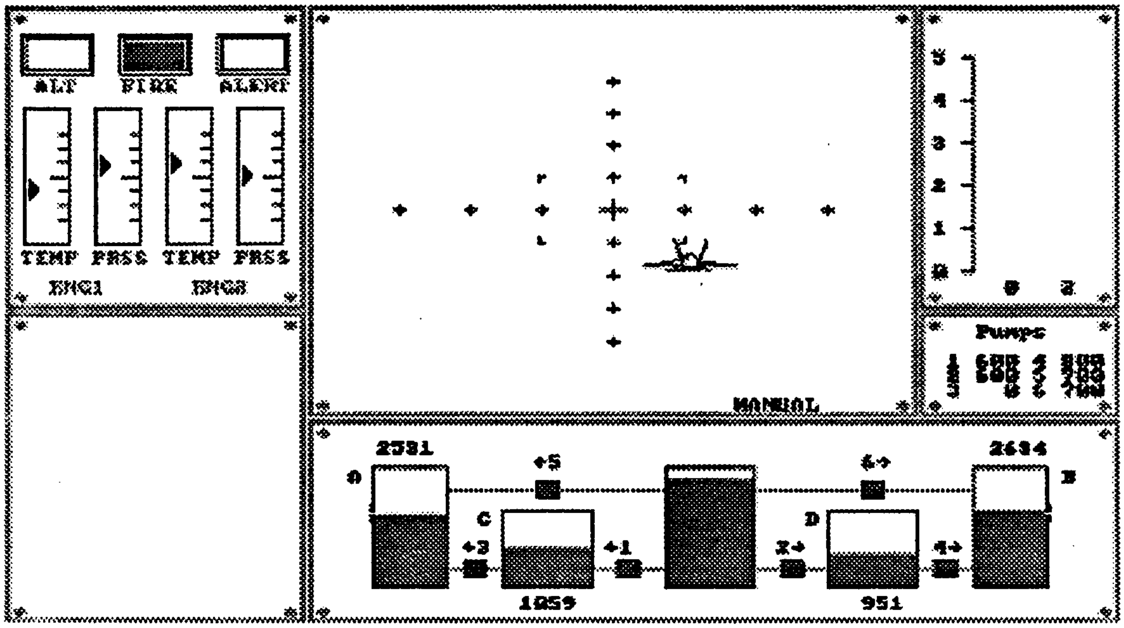


Figure 1. The MINUTES interface with the three major sub-tasks illustrated in detail.

Subjective Workload and Fatigue Assessment

Subjective workload was assessed by the Subjective Work Assessment Technique (SWAT) (Reid & Nygren, 1988). Subjective fatigue was assessed by the Profile of Mood States (McNair, Lorr, & Droppleman, 1971) and the Positive and Negative Affect Scale (Watson, Clark, & Tellegen, 1988). The Profile of Mood States (POMS) provides a self-report of six psychological states, tension, depression, anger, vigor, fatigue and confusion. The estimate of fatigue used a combination of the fatigue and vigor scores of the POMS where Total Energy = vigor - fatigue. The positive and Negative Affect Scale (PANAS) provides estimates of positive and negative mood. The PANAS was administered by placing items in the message window and participants indicated their responses on the keyboard. Critical fusion frequency (CFF) was also assessed using the Pocket Flicker (Saito, & Hosokawa, 1989).

Experimental Procedure

Participants received instructions and practice on each MINUTES sub-task and were provided a one hour practice session. Immediately following completion of the practice session, participants received training on the SWAT. Initially participants completed the POMS, indicated when they could detect pulsation of the Pocket Flicker light source five times, and then began a one hour and forty minute MINUTES session. Participants were randomly assigned to one of two groups. All participants manually controlled tracking during minutes 0 to 20, 40 to 60 and 80 to 100. During minutes 20 to 40 and 60 to 80, participants were given the opportunity to switch between manual and automatic tracking by pushing a button on the joystick. These periods are refereed to as automation periods one and two. During the first automation period for group 1 and the second period for group 2, a bar graph in the upper-right hand corner of the screen indicated current task load and task load for the next two minutes (taskload projection). In the other automation-available period, participants could switch between manual and automatic control but they were provided no information regarding current or projected task load. Following the session, each participant completed the POMS a second time and completed 5 flicker estimation trials.

Resource management and monitoring were required during all 6 phases of the 100 minute MINUTES session. Task load among phases was varied by changing the frequency of lights or gauges that required resetting and the pump failure frequency. Performance assessment included monitoring response time, false alarms, and misses (no response to a signal), and resource management error (sum of absolute errors from 2500 for both tanks). The PANAS was administered 5 minutes after the session began and 5 minutes before the end of the session. The SWAT was administered at 10 minute intervals during the session by placing items in the screen message window.

RESULTS

Automation Invocation

A key question examined here is the use of automation and the effect of expected load preview and fatigue on automation invocation behavior. With respect to the former, we see a reduction in the use of automation from the first to the second period when it is available. Average use dropped from over four and one half minutes (280s) to under three minutes (168s) of the twenty possible minutes available. However, this difference did not reach traditional levels of statistical significance because of the large inter-subject variability, a theme to which we return. With respect to load preview, we see a similar pattern. Without such a facility, the average use of optional automation was over four minutes in length (252s). However, with previews such use dropped to below three and one half minutes (197s). Again, this difference did not reach standard statistical significance as the variability between subjects was high. The factor of individual differences in people's automation strategy is one to which we return in discussion.

	Period 1	Period 2	Projections	No projections
Automation Time (sec)	280	168	197	252
Monitoring RT (Sec)				
Red light	0.66	1.19	0.89	0.96
Green light	2.22	2.94	2.60	2.56
Gauges	8.75	8.12	8.20	8.66
Engine problem	3.18	2.93	2.88	3.22
Missed Signals (Responses)				
Gauges	14.1	32.9	16.75	29.8
Engine problem	4.0	4.5	3.4	5.1
False Alarms (Responses)				
Gauges	7.04	16.44	8.62	14.9
Engine problem	2.00	2.38	1.69	2.69
Fuel Management Error (Gal)	522	1178	771	928

Table 1. Mean performance as a function of automation period and taskload projection availability.

Performance Efficiency

Participants automated tracking time for varying portions of each period and therefore no attempt is made to compare tracking error during the two taskload projection conditions. Fuel management accuracy varied as a function of the time on task. Fuel management was more accurate during the period one than during the second period fuel management, $F(1, 6) = 20.77$, $p < 0.004$. Average resource management error was 522 gallons during period one and 1178 gallons during optional automation period two (see Table 1 below).

Red light response time (RT) increased during period two, $F(1,6) = 13.45$, $p<0.01$, and the gauge false alarm rate increase approached significance. Fuel management error, monitoring RTs and gauge and engine problem missed signals and false alarms increased when projections were not available, however, the increases were not significant. A summary of monitoring response times, missed signals, and false alarm frequency are presented in Table 1. Comparison of the RT to the four monitoring tasks indicated that the response time to the red and green lights and the scales conformed to patterns observed in previous research (Harris, Hancock, Arthur, & Caird, 1991). No failures to respond to the red and green lights, or to the engine problem were observed.

Mental Workload

The SWAT was administered at ten minute intervals during the session. The subjective workload of participants receiving projections did not differ from no projection participants during the first period but the period 2 subjective workload of participants who changed from projections to no projections increased while participants who changed from no projections to projections indicated decreased workload ($p<0.048$).

Subjective Fatigue

Fatigue was estimated by two self-report instruments, the Profile of Mood States (POMS) and the PANAS, and by the Flicker Meter which assessed critical flicker fusion frequency (CFF). Fatigue was assessed by combining two POMS scales, vigor and fatigue, into a total energy score. Total Energy was calculated by subtracting the fatigue scale from the vigor scale. The pre-session total energy score of -1.5 indicated that participants were consistent with the value predicted by college norms but below the levels observed in elite athletes (Morgan, 1985). Participants exhibited a significant total energy decrease during the 100 minute session ($t\{13\} = -2.11$ $p<0.027$). The second psychological state assessment, the PANAS, indicated a stable negative mood and a significant decrease in positive mood during the session ($p<0.028$).

DISCUSSION

When a human operator and a machine cooperatively manage a system, maximum system performance requires efficient interchange of control. When automation begins during periods of low taskload, the workload change is not significant. When taskload increases rapidly and exceeds operator capability, initiating automation threatens to increase taskload at precisely the moment when taskload reduction is needed. The ability to effectively switch between automation and manual control is necessary to realize the full benefits. The present experiment indicates that fatigue decreases participant's ability to use optional automation to manage a high workload, multi-task environment.

The presence of increased mental fatigue during the session was confirmed by the POMS and the PANAS which both indicted increased subjective fatigue during the experimental session. However, CFF decrease reported to be associated with fatigue was not observed. Increased fatigue during high cognitive workload is consistent with previous reports in experimental settings (Harris, Hancock, Arthur, & Caird, 1991), simulators (Harris & Clubb, 1991) and field settings (Rosa & Colligan, 1988). The lack of agreement between self-reported psychological state and CFF changes raises questions regarding the linkage between physiological measures and subjective measures of fatigue. The agreement between the sixty-five item pencil and paper POMS and the on-screen PANAS assessment during the session supports the feasibility of on-line fatigue assessment in multi-task batteries and suggests that psychological state can be assessed without stopping synthetic task performance.

The current study suggests that pilot initiated automation may be less effective when pilots are fatigued and suggests that exploration of the value of taskload projections is worthy of further exploration. Pilot

control has been reported to increase situational awareness and provide a feeling of system control which is compatible with low workload. Increased fatigue would be expected to decrease pilot's use of optional automation and their ability to use available performance aids. Automation initiated on the basis of triggers such as performance, taskload, resource allocation estimates, and/or the nature of task and operator requests for aid (Andes & Rouse, 1991) would appear to be indicated when pilots become fatigued.

In optional automation period one, high resource management error was not observed and participants with the greatest resource management error used automation more frequently. However, during the second optional automation period participants exhibiting high resource management error used automation infrequently. Fatigued operators who were not managing fuel effectively did not use available automation. The value of optional, operator initiated automation is reduced if individuals fail to use available performance aides during periods of high subjective workload, fatigue, and performance deterioration.

The effect of taskload projections on subjective workload appear to be sensitive to the order of presentation and possibly to the level of fatigue. Participants who changed from the taskload projection condition to the no projection condition indicated they were experiencing increased workload but participants shifting from no projection to projection indicated that they were experiencing less workload. Perceived workload thus appears to be sensitive to changes in the availability of taskload information.

Although the instructions to participants were designed to standardize their distribution of resources and develop a consistent pattern of sub-task weighing, examination of the data indicated that the participants took different approaches to managing the many options available in the MINUTES task. The use of automation and the allocation of mental resources to sub-tasks in a multi-task environment were characterized by large individual differences.

In conclusion, a pattern of less frequent use of automation by fatigued participants with performance deterioration suggests that fatigue decreases the effectiveness of participant enabled automation. This observation is consistent with observations that cognitive performance deficits accompany fatigue (Bonnet, 1980) and suggest that automation research should attempt to evaluate such operator states. The willingness to use available performance aids exhibited large individual differences thus it can't be assumed that all individuals use performance aids in the same manner. We conclude as we began. The accumulated evidence reinforces our assertion that just because you can automate a task doesn't necessarily mean that you should.

REFERENCES

Andes, R.C., & Rouse, W.B. (1991). *Specification of adaptive aiding systems.* Search Technology Technical Report. 4725 Peachtree Corners Circle, Suite 200, Norcross, GA.

Bonnet, M.H. (1980). Sleep, performance and mood after the energy-expenditure equivalent of 40 hours of sleep deprivation. *Psychophysiology*, 17, 56-63.

Comstock, J.R., & Arnegard, R.J. (1990). *Multi-attribute task battery.* NASA Langley Research Center, Langley, VA.

Hancock, P.A. (1992). The future of hybrid human-machine systems. In: J.A. Wise, V.D. Hopkin., & P. Stager, (Eds.) *Verification and validation of complex human-machine systems.* NATO, New York: Springer.

Hancock, P.A., & Williams, G. (1993). *Effect of taskload and task load increment on performance and workload.* Proceedings of the conference on aviation psychology, Columbus, OH.

Harris, W.C., & Clubb, J.M. (1991). *An exploratory examination of the mood of Naval aviators,* Paper presented at the Convention of the American Psychological Association.

Harris, W.C., Hancock, P.A., Arthur, E.J., & Caird, J. (1991). Automation influences on performance, workload and fatigue. Paper presented at the Annual Meetings of the Human Factors Society, San Francisco, CA..

Harris, W.C., Hancock, P.A., Arthur, E., & Manning C. (1992). *Minnesota universal task evaluation*

system [MINUTES] . University of Minnesota, Human Factors Research Laboratory Technical Report HFRL 92-02, Minneapolis, MN.

McNair, D.M., Lorr, M., & Droppleman, L.F. (1971). *Profile of mood states manual.* San Diego, CA: Educational and Industrial Testing Service.

Morgan, W.P. (1985). Selected psychological factors limiting performance: A mental health model. In D.H.. Clark & H.M. Eckert (Eds.) *Limits of human performance*, Champaign, IL. Human Kinetics.

Parasuraman, R. Bahri, T., Deaton, J.E., Morrison, J.G., & Barnes, M. (1990). *Theory and design of adaptive automation in aviation systems.* Warminster, PA: Naval Air Development Center (Code 6021)

Reid, G.B., & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock, and N. Meshkati (Eds.), *Human mental workload,* Amsterdam, The Netherlands, Elsevier.

Rosa, R.R., & Colligan, M.J. (1988). Long workdays versus restdays: Assessing fatigue and alertness with a portable performance battery. *Human Factors*, **30**, 305-317.

Saito, K., & Hosokawa, T. W. (1989). A new apparatus for evaluating fatigue of VDT workers, In: M.J. Smith and G. Salvendy (Eds.) *Work with computers: Organizational, management, stress and health aspects*. (pp. 654-658), Amsterdam. Elsevier Science Publisher B.V.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, **54**, 1063-1070.

Weiner, E.L., & Curry, R.E. (1980) *Flight-deck automation: Promises and problems.* Moffett Field, CA: NASA Technical Memorandum 81206.

Weiner, E.L., & Nagel, D.C. (1988). *Human factors in aviation.* San Diego, CA: Academic Press.

This research was supported by NASA Ames Research Center through Grant NAG 2-749 and the Naval Air Warfare Center through contract N/N62269-92-C-0211. Sandra Hart was the NASA Grant Technical Monitor and Jeff Morrison and John Gluckman were the NAWC contract monitors. The views expressed are those of the authors and do not necessarily represent those of the named agencies.

OPERATOR VERSUS COMPUTER CONTROL OF ADAPTIVE AUTOMATION

Brian Hilburn, Robert Molloy, Dick Wong, and Raja Parasuraman
Cognitive Science Laboratory
The Catholic University of America
Washington, DC

INTRODUCTION

Adaptive automation refers to real-time allocation of functions between the human operator and automated subsystems[1]. It has been proposed for some time that automation that is implemented dynamically, in response to changing task demands placed upon the operator, can permit the chief benefits of automation (e.g., workload regulation) to be realized in the aviation cockpit, without some of the drawbacks associated with so-called static automation. One of the chief assumptions underlying the use of adaptive automation is that the pilot (or generally, any operator) can control a process during periods of moderate workload, and hand off control of particular tasks when workload either rises above, or falls below, some optimal level. The issue of how the system should infer workload changes has led to the description of four broad methods for triggering adaptation. These are: pilot *performance measurement*, *psychophysiological assessment*, *performance modeling*, and *critical-events logic* (Parasuraman et al., 1990; Rouse, 1988).

Using a measurement approach, the decision to automate is based upon dynamic assessment of pilot workload, typically using either physiological or behavioral measures. Modeling approaches to adaptation would invoke automation on the basis of impending performance degradation, as predicted by some human performance model. Many such models exist, and can be classified broadly as either optimal performance models (such as signal detection, information and control theories), or information-processing models (such as multiple resource theory). A modeling approach based on, say, multiple resource theory would predict performance degradation whenever concurrent tasks placed excessive demands on common resources. Based on inputs primarily external to the pilot, such a scheme would then decide to invoke automation. Another method for control of adaptive automation, critical-events logic (Barnes & Grossman, 1985), bases adaptation on mission goals. Critical-events logic is in some respects the least technically-difficult scheme to implement in real settings. For instance, if some pre-defined "critical-event" (e.g., sudden appearance of a hostile aircraft) occurs, certain defensive measures are carried out by automation.

There are benefits and drawbacks associated with each of these adaptation methods. Although critical-events logic might be appropriate under emergency circumstances, it fails to consider the actual workload or performance of the operator. Measurement of the operator's mental state (e.g., workload, strategies, vigilance) can, in principle, be carried out on-line, and so offers some promise of flexibly responding to unpredictable changes in pilot cognitive states. Unfortunately, this scheme is only as good as the sensitivity, diagnosticity, and validity of the measures used to trigger adaptation. As an off-line technique, modeling approaches have the advantage that they can be easily incorporated into rule-based expert systems. They have the attendant problem, however, that it is often difficult, particularly in a complex multi-task environment, to adequately specify a priori all eventualities that might be faced in real settings.

It has been proposed that a hybrid system incorporating more than one of these methods might optimize their relative benefits and minimize their drawbacks (Parasuraman et al., 1990). A system combining, for instance, measurement and critical-events logic, or measurement and modeling, might afford a system that optimizes criteria such as operator acceptance, timely function allocation, sensitivity, and robustness. In a related proposal, Corso (1991) suggested that human outputs (i.e., performance or workload

---

[1] in addition to adaptive *function allocation*, several other schemes exist for achieving adaptive automation. For instance, tasks can be adaptively *partitioned*, with human and machine each responsible for some portion of the task.

measurements) should be used to train a system, but that human inputs (i.e., critical-events) should trigger adaptation in the operational setting.

COMPUTER VERSUS OPERATOR CONTROL OF ADAPTATION

One of the fundamental concerns in the design of adaptive function allocation systems has been the relative authority that the operator and the automation should have over the operation of the system. For instance, how should the automation be invoked-- should the human operator or the system have the authority to control changes in the level of automation?

The relative authority that the system should exercise over the invocation of automation can be viewed from a number of perspectives. Barnes and Grossman (1985), recognizing the potential inflexibility in the critical-events method, distinguished three types of logic within this general scheme: *executive logic,* in which the final authority to automate rests with the pilot; *emergency logic,* in which a control process is executed without pilot initiation; and *automated display logic,* in which the system is free to automate all non-critical functions. Notice that this last logic is in fact a form of emergency logic associated with some subset of tasks.

The issue of authority (executive versus emergency) that has been discussed in the context of critical-events logic can be considered separately from the choice of adaptation scheme. For instance, either a measurement or modeling approach to adaptation can be coupled with either type of logic. Such a taxonomy brings the discussion more into line with the view that control of a complex system can vary along a continuum from fully manual to fully automated (McDaniel, 1988; Morris, Rouse & Ward, 1985). Along this continuum, there can be stages at which the human initiates (and the system consents) to adaptation, or vice versa. For instance, the human might be free to request assistance whenever it is desired. Or the system might recommend automation, but surrender final authority to the operator. McDaniel (1988) notes the need to retain at least informed operator consent in cases of especially critical functions, such as weapons launch. Finally, the system might invoke automation unless specifically overridden.

Many approaches to adaptive aiding implicitly assume that operator control of function allocation is preferable to system control, or at the least that operator *consent* to any suggested changes should be mandatory. For example, such a position is consistent with the approach to automated aiding followed in the Pilot's Associate program. Yet, in common with a number of other issues pertaining to adaptive systems, there is little or no empirical evidence by which one might evaluate such a position. In this article we report the results of a series of experiments whose aim is to examine the effects of adaptive automation on operator performance during multi-task flight simulation, and to provide an empirical basis for evaluations of different forms of adaptive logic. Five experiments using the Multi-Attribute Task (MAT) battery are reported. The MAT is a PC-based laboratory flight simulator comprising component tasks of compensatory tracking, system monitoring, and fuel management (Comstock & Arnegard, 1992). The first two studies used an implicit performance modeling logic, whereas studies 3 through 5 experimentally manipulated the type of logic used to invoke automation.

EXPERIMENT 1: COSTS AND BENEFITS OF SHORT-CYCLE ADAPTIVE AUTOMATION

Adaptive automation involves transitions between automated and manual control. This first experiment investigated the costs and benefits of so-called *short-cycle* automation, in which flight functions are cycled between manual and automated control fairly frequently (Parasuraman et al., 1991a). The aim of the experiment was to examine the benefits and possible costs of such dynamic shifts in automation.

Twenty four non-pilots were tested on the MAT battery of compensatory tracking, systems monitoring and fuel management tasks. Tasks could be performed under either manual or automated control. After initial manual practice, subjects each performed four 30-minute sessions. Each session consisted of three 10 minute blocks: manual control (M), automated control (A), and a second manual block, referred to as "return-to-manual" (RM) block. The automated task was varied between subjects, so that a

given subject had one task (e.g., system monitoring) automated during the [A] blocks, while the other two tasks were under automated control. During [M] and [RM] blocks, all three tasks were performed manually.

Automation benefits (M versus A) and costs (M versus RM) were analyzed through separate ANOVAs, for each of the four dependent measures (monitoring RT and accuracy, tracking RMS error, and fuel RMS error). Overall, performance was enhanced by automation of the other two tasks (relative to the preceding manual block). To compare the effect of automating each of the three tasks on the two non-automated tasks, performance on the non-automated tasks was converted to a z score composite. This measure revealed that, across non-automated tasks, automation of the tracking, monitoring and fuel management tasks was associated with a performance increase of .3, .27, and .49 z score units, respectively. This pattern diminished with practice, although automation benefits were observed in each of the four blocks.

No evidence of automation costs was obtained. Although practice effects were found across the four blocks, there were no significant differences within blocks, between the M and RM conditions. In fact, across all measures, the RM condition was associated with a mean performance *improvement* of 3.75%. These results confirmed that adaptive automation can enhance performance, across the tasks studied. While these results were encouraging, it remained to be seen whether such benefits were diminished, or whether costs appeared, with the use of *long-cycle* adaptation, in which functions are automated for extended periods of time.

EXPERIMENT 2: LONG-CYCLE AUTOMATION

One of the dangers of extended periods of automated operation is the increased demand placed on the operator to monitor for potential automation malfunctions. This situation introduces the potential for several human performance problems. First is the possibility that the operator will place excessive trust in the automation. The concept of "complacency" has been used to describe this situation (Parasuraman, Molloy, & Singh, 1993; Wiener, 1981). Second, several authors have suggested that degradation of manual skills, which can accompany extended periods of automation, might limit the pilot's ability to revert to manual control. This is a situation for which humans are not well-suited (Parasuraman, 1987), and has led some to speculate that monitoring tasks are the most likely candidates for computer aiding (Johannsen , Pfendler, & Stein, 1976). This second study investigated these possible costs of automation (Parasuraman et al., 1993).

This study used a similar four session design, with twelve 10-minute blocks in all. Subjects performed the tracking and fuel management tasks of the MAT battery under manual control, with the system monitoring task under partial automation control. That is, the automation responsible for overseeing the monitoring task had programmed "failures" to detect system malfunctions. The subject was responsible for backing up the automation.

The chief result from this study, for the purposes of the present discussion, was that monitoring was relatively inefficient , falling to about 32% overall, even though <u>manual</u> monitoring was quite good (75%). This effect was observed after only 20 minutes of automated control. This was the first empirical evidence that periods of extended static automation might introduce performance problems.

EXPERIMENT 3: THE USE OF ADAPTIVE AUTOMATION TO ENHANCE MONITORING PERFORMANCE

Together, experiments 1 and 2 suggest that there might be some loss of monitoring skill under long-cycle adaptation, that is not apparent under a shorter adaptation cycle. This study (Parasuraman et al., 1992) investigated the hypothesis that adaptive automation could be used to counter the monitoring decrement seen in long-cycle automation interludes. We hypothesized that, following an extended automated session, a reversion to manual control would enhance monitoring performance; further, we expected that the enhancement would persist for some time into the next automated period. A similar design was used for this study, with subjects performing in four 30 minute sessions, resulting in a total of 12 ten-minute blocks. The fuel management and tracking tasks of the MAT were always under manual control.

Control subjects performed all twelve blocks with the system monitoring task under automated control, while experimental subjects faced possible reversion to manual system monitoring during block 5. As before, during automated interludes subjects were responsible for backing up the unreliable automation. Subjects were assigned to one of two adaptive logic groups: A *model-based* adaptive group always reverted to manual monitoring during block 5; *performance-based* subjects reverted during block 5 only if their previous monitoring accuracy fell below a criterion (57%).

The two adaptation groups performed much better under manual control (block 5) than under the pre-allocation automation period (blocks 1-4), relative to the non adaptive control group. This is not surprising, and is consistent with the findings of Wickens and Kessel (1981), who demonstrated that an "in-the-loop" operator monitors better. What is notable, though, was that post-allocation monitoring performance was much better for the two adaptive groups. In block six, the first re-automated block, the adaptive groups detected an average of 60% of system malfunctions, compared to roughly 30% for the non-adaptive control group. This difference diminished over the remaining blocks, suggesting a monitoring decrement similar to that of the pre-allocation phase. Together, these data suggest that an occasional reversion to manual control can enhance subsequent monitoring under automation.

EXPERIMENT 4: OPERATOR VERSUS COMPUTER ADAPTATION

It is often assumed that operator control of function allocation is superior to system control, or at the least that human consent should be retained whenever possible. Thus far, however, there had been no empirical evidence to either support or reject this claim. This study compared performance under two alternate forms of critical-events logic, executive and emergency logic, within a simulated hybrid adaptive system combining these logics with a model-based approach.

One difficulty in trying to compare operator and computer adaptation is experimentally controlling for such confounding effects as (1) subjects' expectancies regarding when the system is going to shift between manual and automated control during computer-controlled sessions, and (2) the frequency with which the system shifts, across the two types of sessions. To control for the former, during computer-controlled sessions an on-screen warning was presented shortly before each automation shift, to alert subjects that control was going to change. To provide some measure of control over the frequency of automation shifts, a complex yoked design was used, in which the pattern of one's subject's executive sessions was used to create the "automation" sequence for a yoked subject's emergency sessions. This controls for both the number and times of automation shifts between the two conditions.

All three tasks of the MAT were again performed. Fuel management and monitoring were always under manual control, while the tracking task could be performed under either automated or manual control. Subjects performed in four 20-minute sessions. During two of these, tracking automation was under computer control, while subjects had control of tracking automation during the other two sessions. Once automation was invoked (by either the subject or the system, depending on the session), the tracking task was performed by completely reliable automation for a fixed two minutes. Once reverted to manual control, the subject was free to re-invoke automation.

Subjects were willing to automate the tracking task (the number of switches per 20-minute session ranged from 3 to 10). Preliminary data analysis revealed several performance trends. First, a composite (z score across measures) score showed a slight automation cost under either form of logic, albeit a smaller cost under executive logic. When performance was plotted as a function of number of switches, however, performance degraded with more automation switches, for each of the three applicable measures. These data suggest that there might be a cost of too frequent cycling between automated and manual control; that is, automation that is excessively short in cycle might also degrade performance. This can represent a potential problem for any adaptive system whose logic is too sensitive, cycling the operator through automated and manual modes at frequent intervals. Second, and more generally, these results indicate that any such performance cost is likely to be reduced (if not eliminated) if automation shifts are under the control of the operator rather than under system control.

EXPERIMENT 5: OPERATOR CONSENT TO AUTOMATION CHANGES

While experiment 4 examined the benefits of operator control in combination with a model-based logic, it was reasonable to speculate that operator control might show differential benefits when combined with model-based or performance-based hybrid systems. Using the basic design of experiment 3, this study examined the usefulness of operator control, when combined with pilot performance measurement.

Subjects were assigned to one of three hybrid adaptation groups: (1) model/emergency, (2) performance/emergency, or (3) performance/executive logic. The monitoring task could be performed under either manual or automated control, while the tracking and fuel management tasks were always performed manually. Subjects in the model group were reverted to manual control during the fifth of nine 10-minute blocks. Subjects in the other two groups remained under automated control if their previous monitoring accuracy remained above 57%. If performance fell below this criterion, however, subjects in the second group (emergency logic) were forced into a manual block, while subjects in the third group (executive logic) were provided the opportunity to override a system reversion to manual control.

Data from this study have not yet been fully analyzed, though initial results suggest a pattern similar to that of experiment 3, in which monitoring benefits were sustained for several blocks following a brief reversion to manual control. No clear advantages of any of the three adaptation conditions have yet emerged, although any final conclusion must await further analyses.

CONCLUSIONS

The combined results of these studies suggest several things. First, it appears that either excessively long, or excessively short, adaptation cycles can limit the effectiveness of adaptive automation in enhancing operator performance of both primary flight and monitoring tasks. Second, occasional brief reversions to manual control can counter some of the monitoring inefficiency typically associated with long cycle automation, and, further, that benefits of such reversions can be sustained for some time after return to automated control. Third, no evidence was found that the benefits of such reversions depend on the adaptive logic by which long-cycle adaptive switches are triggered.

Though not explicitly part of the original design, experiments 3 through 5 compared the effectiveness of model-and performance-based adaptive schemes, as well as the two critical event logics. None of these studies, however, specifically addressed the possibility that benefits might accrue from an adaptive scheme in which the type of adaptive logic depends on whether some measure of recent performance falls either above, or below, some criterion. For instance, at high workload, the choice of adaptation might be best left to machine, while at low workload, the human might be a better judge of the need for automation. It is below the threshold of unimpaired performance that subjective and performance-based measures typically dissociate (Eggemeier & O'Donnell, 1986). Under conditions of underload, subjective reports are generally more sensitive than measures of overt performance. Future studies might address whether there is any advantage to a hybrid system in which the source of adaptation (human or computer) is tied to some measure of current workload or performance.

ACKNOWLEDGEMENTS

Supported by contract N62269-91-C-0240 from the Naval Air Warfare Center (NAWC), Warminster, PA (Jonathan Gluckman, technical monitor), and by research grant NAG-1-1296 from the National Aeronautics and Space Administration (NASA), Langley Research Center, Hampton, VA (Alan Pope, technical monitor). The views presented in this paper are those of the author and not necessarily representative of NAWC or NASA.

REFERENCES

Barnes, M. & Grossman, J. (1985). *The intelligent assistant concept for electronic warfare systems.* (Technical Report NWC TP 5585). China Lake, CA: Naval Warfare Center.

Comstock, J.R. & Arnegard, R.J. (1992). *Multi-attribute task battery.* (Technical Memorandum 104174). Hampton, VA: NASA Langley Research Center.

Corso, G. (1991). *Adaptive function allocation: a review of the human factors issues.* Unpublished manuscript, Georgia Institute of Technology.

Eggemeier, F.T., & O'Donnell, R., (1986). Workload assessment technology. In K. Boff, L. Kaufman, & J. Thomas (Eds.) *Handbook of perception and human performance.* New York: Wiley.

Johannsen, G., Pfendler, C., & Stein, W. (1976). Human performance and workload in simulated landing-approaches with auto-pilot failures. In T. Sheridan & G. Johannsen [Eds.] *Monitoring Behavior and Supervisory Control.* New York: Plenum.

McDaniel, J. (1988). Rules for fighter cockpit automation. Paper presented at the IEEE National Aerospace and Electronics Conference. New York: IEEE, 831-838.

Morris, N.M., Rouse, W.B., & Ward, S.L. (1985). Experimental evaluation of adaptive task allocation in an aerial search environment. *Proceedings of the IFAC Conference on Analysis, Design, and Evaluation of Man-Machine Systems*, 67-72, Varese, Italy.

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors, 29 (6)*, 695-706.

Parasuraman, R., Bahri, T., Deaton, J., Morrison, J., & Barnes, M. (1990). *Theory and design of adaptive automation in aviation systems.* (Technical Report). Warminster, PA: Naval Air Development Center.

Parasuraman, R., Hilburn, B., Molloy, R., & Singh, I.L. (1991a). *Adaptive automation and human performance: III. Effects of practice on the benefits and costs of automation shifts.* (Technical Report CSL-N91-3). Washington, DC: Cognitive Science Laboratory, Catholic University of America.

Parasuraman, R., Bahri, T., Molloy, R., & Singh, I.L. (1991b). *Adaptive automation and human performance: II. Effects of shifts in the level of automation on operator performance.* (Technical Report CSL-N91-2). Washington, DC: Cognitive Science Laboratory, Catholic University of America.

Parasuraman, R., Molloy, R. & Singh, I.L. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology, 3(1)*, 1-23.

Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1992). *Training and adaptive automation: II. Adaptive manual training.* (Technical Report CSL-N92-2). Washington, DC: Cognitive Science Laboratory, Catholic University of America.

Rouse, W.B. (1988). Adaptive aiding for human/computer control. *Human Factors, 30*, 431-443.

Thackray, R.I. & Touchstone, R.M. (1989). Detection efficiency on an air traffic control monitoring task with and without computer aiding. *Aviation, Space, and Environmental Medicine, 60*, 744-748.

Wickens, C. D. & Kessel, C. (1981). Failure detection in dynamic systems. In J. Rasmussen and W.B. Rouse (Eds.), *Human detection and diagnosis of system failures.* New York: Plenum.

Wiener, E. (1988). Cockpit Automation. In E.L. Wiener & D.C. Nagel (Eds.) *Human Factors in aviation.* San Diego: Academic Press.

ADAPTIVE FUNCTION ALLOCATION REDUCES PERFORMANCE COSTS OF STATIC AUTOMATION

Raja Parasuraman, Mustapha Mouloua, Robert Molloy, & Brian Hilburn
Cognitive Science Laboratory, The Catholic University of America
Washington DC

INTRODUCTION

Adaptive automation offers the option of flexible function allocation between the pilot and on-board computer systems. The basic concept was proposed some time ago (e.g., Rouse, 1977). However, adaptive automation has only recently been considered to be a viable design option for the cockpit (Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1990; Rouse, 1988). Conventional or "static" cockpit automation has produced many system benefits. At the same time, some problems have arisen, such as reduced system awareness and manual skills degradation (Wiener, 1988). Systems in which automated aids are implemented dynamically, in response to changing system demands, are thought to be less vulnerable to such problems because they provide for regulation of workload, maintenance of skill levels, and task involvement (Hancock & Chignell, 1988; Parasuraman et al., 1990; Rouse, 1988; Wickens, 1992).

Thus far, these proposals remain mere claims. Empirical tests of their validity have only just begun (Parasuraman, 1993). The advantages and possible drawbacks of adaptive function allocation need to be examined for a broad range of flight functions. Parasuraman et al. (1990) identified four major procedures for implementing adaptive automation in the cockpit: (1) critical events; (2) pilot performance measurement; (3) pilot psychophysiological assessment; and (4) pilot modeling. The theoretical benefits and disadvantages of each of these methods of adaptation have been discussed (Parasuraman et al., 1990; Rouse, 1988). Irrespective of the relative merits of these procedures, however, the basic question of the general effectiveness of adaptive function allocation remains to be addressed comprehensively.

One of the important claims for the superiority of adaptive over static automation is that such systems do not suffer from some of the drawbacks associated with conventional (nonadaptive) function allocation. Several experiments designed to test this claim are reported in this article. The efficacy of adaptive function allocation was examined using a laboratory flight-simulation task involving multiple functions of tracking, fuel-management, and systems monitoring (Comstock & Arnegard, 1992).

MONITORING OF AUTOMATION FAILURES

Parasuraman, Molloy, and Singh (1993) showed that operator detection of automation failures is substantially degraded in systems with static automation in which function allocation between operator and system remains fixed over time. Nonpilot subjects performed tasks of tracking, fuel-management, and systems-

monitoring task for several 30-min. sessions (each consisting of three 10-min. blocks). Subjects performed tracking and fuel-management tasks manually, while the systems-monitoring task was performed under automation control. However, subjects were required to detect automation "failures" by identifying engine malfunctions not detected by the automation. Although subjects could easily detect these malfunctions when they did the task manually, under automation, the detection rate was substantially degraded, especially when other manual tasks had to be performed. The mean detection rate of automation failures was only about 32% even though subjects detected over 75% of malfunctions in the manual condition. This effect was apparent after about 20 minutes spent under automation control. These results provide a clear indication of the potential cost of long-term static automation on system performance.

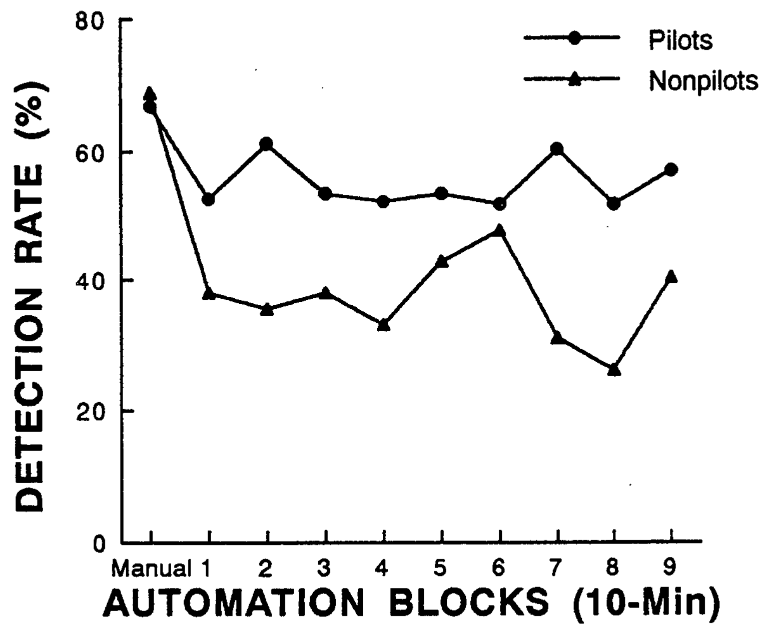


Figure 1. Monitoring of automation failures by pilots and nonpilots.

Experienced pilots show similar performance trends. Figure 1 compares the performance of 8 pilots and 12 nonpilots. As Figure 1 shows, while pilots detected about 70% of malfunctions under manual control, the detection rate dropped to about 55% in the automation blocks. Although the overall performance level of the pilots was higher than that of the nonpilots, the pilots showed the same pattern of performance decrement under automation as did the nonpilots.

ADAPTIVE FUNCTION ALLOCATION AS A COUNTERMEASURE TO MONITORING INEFFICIENCY

Given that both nonpilots and pilots are relatively inefficient in monitoring automation failures for a task that is automated for long periods of time, a logical next step was to examine whether adaptive task allocation provides a counter-measure. The effects of brief periods of manual task allocation on subsequent operator monitoring of the task under automation were examined. Eighteen

nonpilot subjects performed the same flight-simulation task for three 30-minute sessions. A control group performed the task under conditions of static automation, as in the previous study. For the *model-based* adaptive function allocation group, a single 10-minute block of fully manual performance on the systems-monitoring task was allocated to subjects in the middle of the second session, i.e., on block 5. This type of function allocation is termed model-based because it reflects a model indicating that operator performance of that function is likely to be inefficient at that point in time (Parasuraman et al., 1990). This method, however, is insensitive to the actual performance of an individual operator. For the *performance-based* adaptive group, function allocation was changed in the middle of the second session for an individual subject only if the past history of that subject's monitoring performance did not meet a criterion. If the criterion was met, the function was not allocated to the subject but continued under automation control. Following 10 minutes of manual performance in block 5, a prewarned re-allocation of the monitoring task back to automation control was initiated. Subjects completed the rest of the second session and the entire third session (blocks 6-9) with automation.

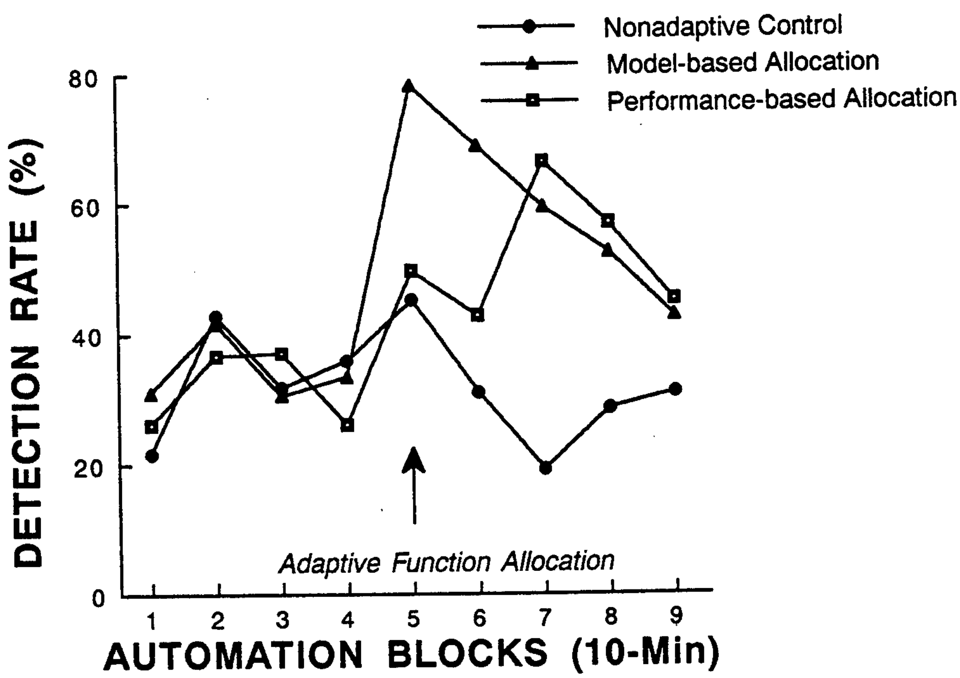


Figure 2. Effects of two methods of adaptive task allocation on detection of automation failures

The results are shown in Figure 2. The detection rate of automation failures was not significantly different for the three groups for the first 40 minutes (blocks 1-4) spent with automation. All groups had low detection rates averaging about 35%. However, detection rates were higher for the adaptive groups than for the control group for the automation blocks following the change in allocation. The overall benefit was approximately the same for both methods of task allocation, averaging about 62% for the two methods.

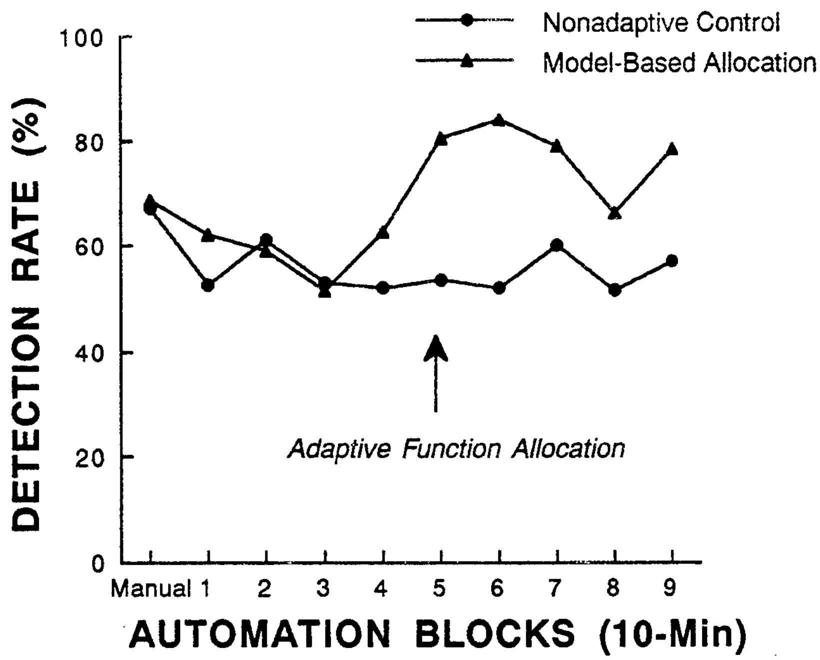


Figure 3. Effects of model-based adaptive task allocation on detection of automation failures by pilots.

Experienced pilots also showed similar benefits of adaptive function allocation, although their overall performance level was higher than that of nonpilots. As Figure 3 shows, performance prior to the function change (Blocks 1-4) was the same in the model-based and control groups. Following the manual block interlude, however, performance was significantly higher for the model-based group than for the control group.

MULTIPLE (REPETITIVE) FUNCTION ALLOCATION

In the previous studies, the performance benefits associated with adaptive task allocation, although relatively large, tended to dissipate over time. The reduction in benefits is understandable given that time may reduce subjects' involvement in the automated task. This raises the issue: can the performance benefits be sustained over longer periods if the automated task is repetitively re-allocated to manual control at multiple time intervals? If empirical evidence for such maintenance of benefits could be obtained the implications would be significant, because it would essentially suggest that performance inefficiency under automation could be eliminated for an indefinite period of time.

A recent study has provided such evidence. A group of nonpilot subjects performed in the same task conditions as the model-based adaptive group in the previous study. However, instead of a single reversion to manual control, subjects were allocated the monitoring task on two occasions, in the middle of the second session (Block 5) and in the middle of the third session (Block 8). For a control group, the monitoring task remained under automation control throughout. In other respects the study was the same as the previous one.

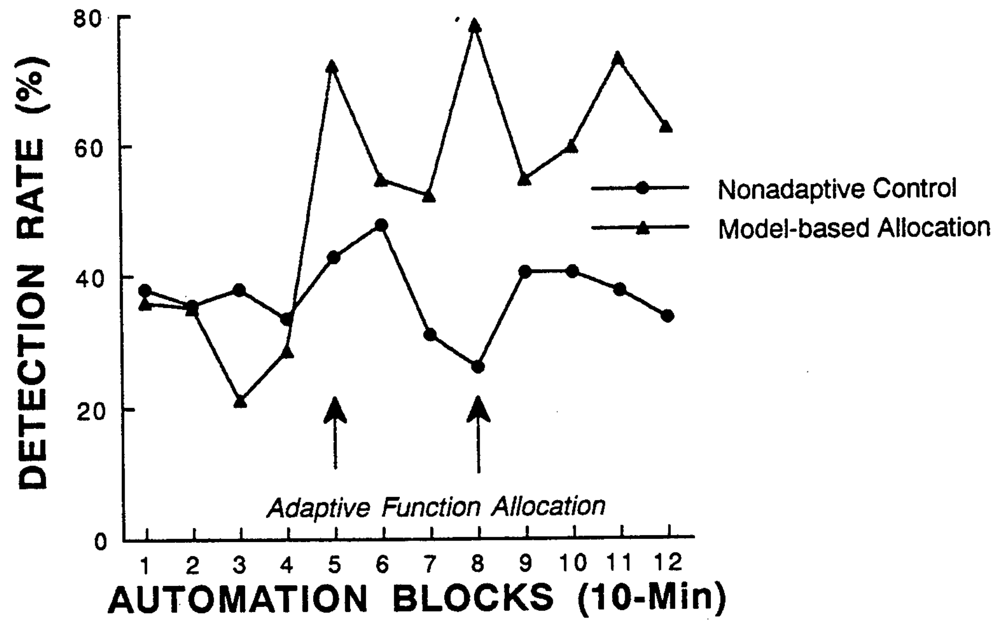


Figure 4. Effects of multiple adaptive function allocation on detection of automation failures.

Figure 4 shows the results. In the four blocks prior to the first function change, both groups had low detection rates, as in the previous studies. Following the function re-allocation in Blocks 5 and 8, however, performance improved, and the overall level of performance in the post-allocation phases was significantly better than that in the pre-allocation phase.

These results indicate that the performance benefits associated with adaptive task allocation can be sustained over prolonged periods of time. The specific schedule of multiple re-allocations, i.e. the time intervals between allocation changes, can be determined given a criterion level of performance that must be achieved: higher performance levels will require more frequent reversions to manual control than lower criterion levels of performance. Similarly, for non-monitoring tasks where automated aiding improves performance, the frequency of allocating the aid will vary directly with the required level of task performance.

CONCLUSIONS

The results of these studies show that monitoring inefficiency represents one of the performance costs of static automation. This cost develops after a fairly short period of time under automation control—20 min. for our flight-simulation tasks. Adaptive function allocation can reduce the performance cost associated with long-term static automation. In our studies, a temporary return to manual control of a previously automated function was found to reduce failures of monitoring, at least for a limited period of time. These effects were observed for both nonpilots and experienced pilots. More sustained benefits were obtained with multiple or repetitive task allocation. Both model-based and performance-based adaptation

produce similar benefits. Choosing between the two methods of adaptation may therefore be need to be based on other criteria, such as perceived workload or pilot preferences (Parasuraman et al., 1990).

The results provide one of the first empirical demonstrations of the efficacy of adaptive task allocation. Most previous reports testifying to the benefits of adaptive automation have either been theoretical, or based purely on anecdotal reports. As Parasuraman et al. (1990) pointed out in their comprehensive survey of this research area, what has been lacking is empirical evidence for (or against) the effectiveness of adaptive function allocation. The present study represents positive evidence with respect to one issue—monitoring of automation failures. Several other issues----training needs, the effects of task allocation versus task partitioning, operator versus system control of allocation decisions, to name a few---remain to be examined systematically.

ACKNOWLEDGMENTS

Supported by contract No. N62269-91-C-0240 from the Naval Air Warfare Center (NAWC), Warminster, PA (Jonathan Gluckman technical monitor), and by research grant NAG-1-1296 from the National Aeronautics and Space Administration (NASA), Langley Research Center, Hampton, VA (Alan Pope, technical monitor). We thank Cdr. John Deaton (NTSC, Orlando, FL) and Richard Gibson and Srinath Sogal (Embry-Riddle Aeronautical University), for help in testing pilots. The views presented in this paper are those of the authors and not necessarily representative of NAWC or NASA.

REFERENCES

Comstock, J. R. & Arnegard, R. J. (1992). *Multi-attribute task battery.* (Technical Memorandum 104174). Hampton, VA: NASA Langley Research Center.

Hancock, P. A., & Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-18,* 647-658.

Parasuraman, R. (1993). Effects of adaptive function allocation on human performance. *Proceedings of the FAA/NASA Conference on Artificial Intelligence and Human Factors in Air-Traffic Control and Aviation Maintenance.* Daytona Beach, FL: Embry-Riddle Aeronautical University.

Parasuraman, R., Bahri, T., Deaton, J., Morrison, J., & Barnes, M. (1990). Theory and design of adaptive automation in aviation systems. (Technical Report). Warminster, PA: Naval Air Development Center.

Parasuraman, R., Molloy, R., & Singh, I.L. (1993). Performance consequences of automation-induced "complacency." *International Journal of Aviation Psychology, 3* , 1-23.

Rouse, W. B. (1977). Human-computer interaction in multitask situations. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-7,* 384-392.

Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human Factors, 30,* 431-443.

Wickens, C. D. (1992). *Engineering psychology and human performance. 2nd edition.* New York:

Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener & D. C. Nagel (Eds.) *Human factors in aviation.* San Diego: Academic Press.

Mr. Christopher Whaley
Georgia Institute of Technology
School of Psychology
Skiles Bldg.
Atlanta, GA 30332

Dr. Robert J. Wheery, Jr.
Chi Systems Inc.
Gwynedd Plaza III
Bethlehem Pike @ Sheble Lane
Spring House, PA 19477

Mr. Jonathan Kern
Veda, Inc.
730 Louis Dr.
Warminster, PA , 18974

Dr. Marion Kibbe
NAWC-WD
Code C2152
China Lake, CA 93555

Dr. Susan S. Kirschenbaum
Naval Undersea Warfare Center
Code 2212, Building 1171/1
Newport, RI 02841

Dr. Scott Makeig
Naval Health Research Center
P.O. Box 85122, Code 233
San Diego, CA 92138-9174

Dr. William F. Moroney
The University of Dayton
Department of Psychology
Human Factors Program
300 College Park
Dayton, OH 45469-1430

Dr. Jeffrey G. Morrison
Naval Air Warfare Center-Aircraft Division
P.O. Box 5152, Code 6021
Warminster, PA 18974-0591
(25 Copies)

Dr. Raja Parasuraman
Catholic University of America
Department of Psychology
Washington, DC 20064
(4 Copies)

Dr. Alan T. Pope
NASA-Langley Research Center
Mail Code 152
Hampton, VA 23665

Mr. Dennis Holeman
SRI International
Mail Stop BN 390
333 Ravenswood Ave.
Menlow Park, CA 94025

Dr. Mark Smolensky
U.S. Department of Transportation
Federal Aviation Administration
FAA Technical Center
ATC Technology Branch
ACD-340
Atlantic City Airport, NJ 08405

(TOTAL: 74 Copies)